This is wrong, right?

Jansma, Dorinde Jennechje

This is Wrong, Right?

The Role of Moral Components in Anti- and Prosocial Behaviour in Primary Education

Dorinde Jansma

This work was funded by the Research Talent Grant (Grant 406-12- 109) of the Dutch Organization for Scientific Research (NWO)

and carried out in the context of the research school of ICO (Interuniversity Center for Educational Research)

ISBN 978-94-034-0262-8

Cover pictures by Emma Wilson
Printed by Ipskamp

rijksuniversiteit
groningen

This is Wrong, Right?

The Role of Moral Components in Anti- and Prosocial Behaviour in Primary Education

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. E. Sterken
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Thursday 25 January 2018 at 16.15 hours

by

Dorinde Jennechje Jansma

born on 9 May 1988
in Dokkum

Supervisors
Prof. M.P.C. van der Werf
Prof. T. Malti

Co-supervisor
Dr. M.C.J.L. Opdenakker

Assessment committee
Prof. D. Brugman
Prof. R.J. Bosker
Prof. R. Veenstra

Table of Contents

Chapter 1
General Introduction

1. Introduction

Antisocial behaviour is an extensive problem in primary education. Children who exhibit antisocial behaviour are not only a great burden for teachers, peers and classroom climate, but also for society, both financially and socially (Soepboer, Veenstra & Verhulst, 2006). Therefore, there is an increasing awareness of the need for educational systems to encourage the acquisition of prosocial values and behaviour, i.e. voluntary behaviour that benefits others (Eisenberg, Spinrad & Knafo, 2015), and to discourage the acquisition of antisocial values and antisocial behaviour, i.e. behaviour that harms others (Brown, Corrigan & Higgins-D'Alessandro, 2012; Fink & Slade, 2016; Rupp & Veugelers, 2003). In practice, this awareness is reflected by numerous educational and intervention programs aimed at affecting bullying, prosocial learning and moral behaviour (Reiman & Dotger, 2008; Smith, Ananiadou & Cowie, 2003; Smith, Cousins & Stewart, 2005). However, research has revealed that the effectiveness of these programs is generally mixed (Wienke, Anthonijsz, Abrahamse, Daamen & Nieuwboer, 2014; Willems, Denessen, Hermans & Vermeer, 2012). Programs directed at social emotional learning show a considerable variety in efficacy (Weare & Nind, 2011). Also, most meta-analyses of anti-bully programs show small to moderate effect sizes at best (Merrell, Gueldner, Ross & Isava, 2008; NJI, 2015). Moreover, positive effects are more likely to be effects on attitudes, knowledge, and perceptions, rather than effects on behaviour (Rivara & Le Menestrel, 2016). Interestingly, some interventions programs do succeed, but not enough is known to indicate exactly how and when (Smith et al., 2004). Beneficial outcomes seem to be distributed across types of assessment, variables and interventions (Merrell et al., 2008). In line with this, Gravemeijer and Kirschner (2007) argue that research aimed at education innovation should not merely focus on evidence of effectiveness (What works?) but instead on understanding the processes explaining the effectiveness (How does it work?).

Thus, research is needed into the fundamental processes underlying anti- and prosocial behaviour. This means that further empirical evidence should validate possible fundamental processes leading to pro- and antisocial behaviour development. By examining the way in which the development of pro- and antisocial behaviour takes place and which factors influence this, (educational) interventions might subsequently be better able to affect both pro- and antisocial behaviour. The current study will focus on moral functioning as a fundamental

process driving anti- and prosocial behaviour in middle childhood (cf. Cuevas, 2011; Caravita, Gini & Pozzoli, 2012; Gasser & Keller, 2009; Gini, 2006; Hymel, Rocke-Henderson & Bonanno, 2005; Menesini & Camodeca, 2008; Perren & Gutzwiller-Helfenfinger, 2012; Pozzoli & Gini, 2010). Moral functioning refers to the psychological process that a person invokes in order to respond to and resolve a specific problem, conflict or dilemma that requires a moral decision and moral behaviour (Tappan, 2006). Both moral decisions and moral behaviour are in turn characterized by "the interest or welfare either of society as a whole or at least of persons other than the judge or agent" (Gewirth, 1984, p. 978). Both prosocial and antisocial behaviour are prime examples of morally relevant behaviour in middle childhood, because of its direct effect on the welfare of others (Fabes, Carlo, Kupanoff & Laible, 1999; Turiel, 1983; 1998). Consequently, moral functioning might serve as a central process underlying children's anti- and prosocial behaviour in primary education.

However, it still remains unclear how different processes of children's moral functioning simultaneously affect anti- and prosocial behaviour (Gasser, Malti & Gutzwiller-Helfenfinger, 2012). Traditionally, research on moral functioning mainly focused on the process of moral reasoning, i.e. justifications for giving a particular moral judgment (Jordan, 2007). However, moral reasoning predicts only 10-15% of the variance in moral behaviour (Blasi, 1980). By now it has been recognized that a comprehensive account of (im)moral behaviour should not only include cognitive processes, but emotional and self-related processes as well (Arsenio & Lemerise, 2004; Olthof, 2010). Rest's (1983; 1986) Four Component Model currently probably offers the most adequate view on morality (Vozzola, 2014). The Four Component Model is grounded in a review of psychological research and is a widely used framework to assess the underlying psychological processes of moral behaviour (Myyry, Juujärvi & Pesso, 2010): moral sensitivity, moral reasoning, moral motivation, and moral character. The Four Component Model is shown in Figure 1.

According to Rest (1986) all moral components, i.e. psychological processes, must be in place in order to act moral. The first component, moral sensitivity concerns interpreting a situation in terms of how people's welfare is affected by possible actions of the subject. The second component, moral reasoning, regards integrating various considerations to determine what ought to be done. Moral motivation, the third component, concerns the importance people give to moral

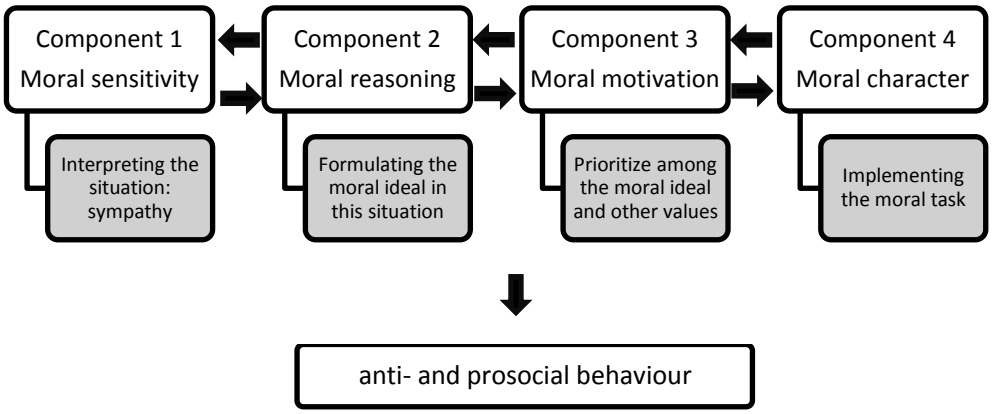


Figure 1. The Four Component Model of Rest (1983; 1986)

values (doing what is right) relative to other values (i.e. self-actualization). The fourth component is moral character and refers to the ability to persist in a moral task in the face of obstacles (Rest, 1983; 1986; 1994).

1.1 Aim of the project and research questions

The aim of this project is to obtain fundamentally new insights into moral functioning and into the relative contribution of the aforementioned four components of Rest's (1983; 1986) model to anti- and prosocial behaviour in primary education. Surprisingly, the relations between these components - moral sensitivity, moral judgment, moral motivation and moral character - and behaviour have remained relatively unexplored in a simultaneous manner (Hardy, 2006; Morton, Worthley, Testerman & Mahoney, 2006). Typically different combinations of only two components have been investigated (Bebeau, 2002), thereby missing the opportunity to simultaneously unpack the processes underlying moral growth and development (Myyry, Juujärvi, Pesso, 2010). Moreover, Rest's model has mainly been studied among adults (Jordan, 2007). Because the seeds of anti- and prosocial behaviour emerge in childhood (Hepach, Vaish, Grossmann & Tomasello, 2016; Malti & Dys, in press), knowledge of moral functioning in these years is essential to understand and intervene in antisocial behaviour and to promote prosocial behaviour. Therefore, this thesis examines all four moral components of Rest's model in relation to anti- and prosocial behaviour among children age 7-12.

2. The four components and anti- and prosocial behaviour

2.1 Operationalization of the four components

The operationalization of the four components has been challenging ever since Rest came up with the Four Component Model (Bebeau, 1994; Myyry et al., 2010). The first component, moral sensitivity, was originally defined by Rest (1986) as a combination of one's recognition of moral issues and how one reacts and processes these issues from an affective perspective within a social context. The current definition now includes dimensions such as 'interpreting others' reactions and feelings', 'having empathy and role-raking ability', 'understanding how one's actions can affect the welfare of both oneself and others', and even 'making inferences from other's behaviour and responding appropriately to their reactions' (Jordan, 2007). Interestingly, no measures exist that directly assess children's moral sensitivity. Instead, several measures assess components of moral sensitivity, such as sympathy and prosocial reasoning (Jordan, 2007). Sympathy is a good representation of moral sensitivity since it constitutes of both empathy and role taking, both very important elements of moral sensitivity (Bebeau, Rest & Narvaez, 1999; Mower, Robinson & Vandenberg, 2015). Sympathy can be defined as feelings of concern or sorrow for the other person based on an understanding of that person's circumstances (Zhou, Valiente & Eisenberg, 2003). It thus concerns all previously mentioned dimensions of moral sensitivity, except for 'making inferences from other's behaviour and responding appropriately to their reactions'. Especially the latter part 'responding appropriately to other's actions' is not captured by sympathy. Nevertheless, one could argue that 'responding appropriately to other's actions' should not be part of the definition of moral sensitivity since it suggests an action instead of a psychological process preceding moral action (e.g. Rest, 1986).

The second moral component, moral reasoning, has been most studied and operationalized of all moral components (Blasi, 1980; Jordan, 2007). This is because for over two decades, research on moral development has been dictated by rationalist models (e.g. Kohlberg, 1969; Piaget, 1965) in which cognition, e.g. moral reasoning, was considered to be responsible for moral action (Rest, 1994). Moral reasoning describes the process in which individuals, using logic and self-reflection, determine why a specific act is right or wrong from a moral perspective (Malti & Ongley, 2014). The dominant stage theory of Kohlberg (1969) stated that there are

hierarchical stages of moral reasoning in which children are first selfish, then oriented to familial and societal regulation, and then formulating principled morality in adolescence. However, more recent moral theories separate moral and non-moral reasoning, allowing for the view that humans prioritize moral over non-moral issues during the lifespan, and not simply as an advanced stage in moral reasoning (Turiel, 1983). Specifically, moral domain theory proposes that moral knowledge (principles of how individuals ought to treat one another) is different from other forms of social knowledge, such as societal knowledge (regulations designed to promote the smooth functioning of social groups and institutions) and psychological knowledge (an understanding of self, others, and beliefs about autonomy and individuality). Thus, beginning in early childhood, children construct moral, societal, and psychological concepts in parallel, rather than in succession. Social domain studies have provided ample evidence that already at 4 years of age, children have developed an understanding of the validity of norms of justice and care, and that they are able to distinguish moral norms from other social rules (Helwig, Tisak, Turiel, 1990; Smetana & Killen, 2008). In this way children are just as well seen as capable moral beings (Nucci, 2001), allowing the investigation of variations in their moral reasoning ability. In correspondence with this view, moral reasoning is nowadays assessed using a distinction between moral (i.e., referring to moral norms and empathic concern for the victim) and non-moral reasoning instead of distinguishing different stages of moral reasoning across development (e.g. Malti, Gasser & Buchmann, 2009; Malti & Ongley, 2014).

Moral motivation, the third component, is conceptualized as the readiness to abide by a moral rule that a person understands to be valid, even if this motivation is in conflict with other, amoral desires and motives (Nunner-Winkler, 1999, 2007). Moral motivation thus implies that a child not only understands, but also personally accepts the validity of moral norms. A central approach to the study of children's moral motivation has been to focus on the anticipation of emotions following moral transgressions, i.e. violations of a moral norm (Nunner-Winkler, 1999; Nunner-Winker & Sodian, 1988). A strength of this approach is that it is a production measure instead of a self-report questionnaire. The underlying assumption of this production measure is that children's emotion attributions represent authentic expressions of what is important to them in a given moral conflict. Thus, moral emotion attributions are seen as indicating the degree to which a child feels personally committed to moral principles and hence also the degree to

which moral principles are integrated into the self (Arsenio et al., 2006; Keller, 1996; Mascolo & Fischer, 2010; Montada, 1993; Nunner-Winkler, 1999, 2007). However, Krettenauer and Malti (2008) debated whether emotion attributions are really an indicator of moral motivation. They claim that changes in emotion attributions are not only a motivational but also a cognitive phenomenon. This means that a child's cognitive capacities, involving mental processes like attention, memory, and thinking, are thought to influence the attribution of emotions In response to this, Nunner-Winkler (2013) argued that positive emotions after a transgression are also present in adults, and that therefore these emotions cannot only be a lack of cognitive capacities and must have a motivational element.

Moral character, the fourth component, is probably the most difficult component to capture by research designs. The vagueness of the moral character component - it includes dimensions like personality traits as well as diverse situational factors - makes it hard to study (Myyvy et al., 2010). Rest (1986, 1994) described moral character as having courage and implementation skills to carry out a line of action in the face of obstacles. The processes that lead to success in this respect are referred to by Rest (1986) as involving ego strength or self-regulation, i.e. choosing for larger delayed rewards or goals for which one must either wait or work. He also noted, however, that ego-strength may be used for the good as well as the bad. Equating moral character to self-regulation therefore does not seem to capture the 'moral' part of moral character and should be complemented with implementation skills or characteristics that help children succeed in doing the good (and not the bad) (Bebeau, 1994). In line with this, the personality characteristics conscientiousness and agreeableness are seen as the precursors of moral character in childhood since they help children succeed in doing the good (Lapsley & Hill, 2009). Moreover, both concepts have a close link to self-regulation. Whereas conscientiousness concerns traits related to self-discipline, orderliness and goal pursuit, agreeableness has a strong link with the regulation of emotions and constitutes of traits related to a desire to maintain social harmony (Ahadi & Rothbart, 1994; Cumberland-Li, Eisenberg & Reiser, 2004; Weisberg, DeYoung & Hirsh, 2011). Additionally, a central self-regulation ability is inhibitory control, i.e. behavioural and cognitive suppression of interferences from the environment (Ahadi & Rothbart, 1994; Graziano & Eisenberg, 1997).

2.2 Relations between the moral components and anti- and prosocial behaviour

Previous research into associations between anti- and prosocial behaviour and Rest's four components rendered various results. Importantly, prosocial and antisocial behaviour are not the opposite ends of a single dimension (Hawley, Little & Card, 2007), and might therefore also differently relate to moral functioning. Whereas prosocial behaviour is voluntary behaviour that benefits others (Eisenberg, Spinrad & Knafo, 2015), antisocial behaviour is behaviour that harms or injures others (Brown, Corrigan & Higgins-D'Alessandro, 2012; Fink & Slade, 2016). A well-known reflection of antisocial behaviour in education is bullying behaviour. Therefore, bullying-related behaviour, and its relation to moral functioning, will receive special attention in this thesis. Bullying occurs in a context of an imbalance of power and is characterized by a repetition of negative actions towards a peer, with the intention to hurt (Olweus, 1993; Veenstra et al., 2005). Moreover, bullying always occurs in a peer group, where other children besides the bully also have a role in influencing the bullying. There are five participant roles involved in the bullying process, next to the victim: bully, who starts the bullying; assistant, who joins in the bullying, but does not start it; reinforcer, who encourages the bully; defender, who supports the victim; and outsider, who keeps out of the bullying situation (Salmivalli, 2010).

The first component, moral sensitivity, expressed as sympathy, negatively correlates with different forms of antisocial behaviour (Björkqvist, Österman & Kaukiainen, 2000; Kaukiainen et al., 1999; Kaukiainen, Björkqvist, Österman & Lagerspetz, 1996; Zahn-Waxler, Cole, Welsh & Fox, 1995) and with bullying in particular (Gini, Pozzoli & Hauser, 2011; Jolliffe & Farrington, 2006). In a review of Van Noorden, Haselager, Cillessen and Bukowski (2015) it was concluded that sympathy showed no association with being victimized and a negative or no association with assisting or reinforcing in the bullying process. Additionally, Sutton, Smith and Swettenham (1999) and Arsenio and Lemerise (2001) found that bullies might not appreciate the emotional consequences of their behaviours on others' feelings and fail to sympathize with the feelings of others. In line with this, Caravita, Di Blaso and Salmivalli (2010) concluded that bullies were less able to understand the pain of other children. On the other hand, they found sympathy to be positively associated with defending. More generally, Hoffman (2000) and Arsenio and Lemerise (2004) argued that children's capacity for sympathy is a key contributor prosocial behaviour. Links between sympathy and prosocial behaviour

have been found to exist both within specific and at the dispositional level (Eisenberg et al., 2015; Malti et al, 2016).

The second component, moral reasoning, was found to explain about 10 - 15% of variance in real life moral behaviour, although empirical support for this association varies from area to area (Blasi, 1980). Generally, moral reasoning is related negatively to delinquency, cheating, aggression and other forms of antisocial behaviours (Eisenberg et al., 2015). Research shows inconsistent findings with regard to the relation between moral reasoning and prosocial behaviour with small to modest positive relations at best (e.g. Eisenberg et al., 2015). Some theorists argue that children's adequate cognitive moral understanding, i.e. moral reasoning, by no means implies they are competent moral actors (Haidt, 2001; Nunner-Winkler, 1999). Simply stated, knowing the good might not be sufficient to do the good (Nucci, 2001). Two studies within the age range of this study empirically demonstrate this proposition. Both Gini, Pozzoli and Hauser (2011) and Olthof (2010) found that bullies know just as well as defenders that some reasons to explain behaviour are morally wrong. On the other hand, Gini (2006) found that 8 to 11 year old bullies, assistants and reinforcers showed less moral reasoning than defenders and outsiders.

The relations between anti- and prosocial behaviour and children's anticipated emotions, expressing the third component moral motivation, have been summarized in a meta-analysis by Malti and Krettenauer (2013). This meta-analysis showed small-size relations between negative emotion attributions and prosocial behaviour and moderate-size relations between negative emotion attributions and antisocial behaviour independent of age. This is in line with theoretical accounts of moral emotions stating that moral emotions may be an important part of why children apply moral justification in complex moral situations and why they adhere or fail to adhere to their own moral standards (Malti & Latzko, 2012; Tangney, Stuewig & Mashek, 2007). Studies of children's anticipated emotions generally assume that children remember the emotional antecedents and consequences of social situations and that this connection between events and emotions guides future behaviour (Arsenio, 2014). Specifically, anticipated negative emotions following moral transgressions restrict aggressive behaviour and motivate reparative behaviour, such as apologies (Asendorpf & Nunner-Winkler, 1992; Malti & Keller, 2010; Tangney et al., 2007). For behaviour in bullying situations, comparable results were found. Gasser and Keller (2009) report that bullies showed

a deficit in moral motivation compared to prosocial children. In the study of Menesini et al. (2003) bullies attributed more negative emotions to a wrongdoer compared to victims and outsiders. Defenders showed less negative emotion attributions than victims.

Moral character, represented by inhibitory control and by the personality dimensions agreeableness and conscientiousness, has been positively related to prosocial behaviour and negatively to antisocial behaviour. Generally, children who have good control over their emotions are more likely to exhibit prosocial behaviours (Beauchaine et al., 2013; Carlo, Crockett, Wolff & Beal, 2012; Laible, Carlo, Panfile, Eye & Parker, 2010; Padilla-Walker & Christensen, 2011). Moreover, the personality of moral exemplars has repeatedly been found to orient toward conscientiousness and agreeableness (Walker, 1999; Walker & Hennig, 2004). Also, Robins, John, and Caspi (1994) have found these two personality dimensions to be strongly and negatively correlated with antisocial personality. In the study of Tani, Greenman, Schneider and Fregoso (2003) defenders exhibited high levels of agreeableness in comparison to other participant roles in the bully process. Contrarily, it has been found that bullies and assistants tend to score lower on agreeableness, conscientiousness and inhibitory control compared to other participant roles (Fossati, Borroni & Maffei, 2012; Menesini et al., 2010; Miller, Lynam & Leukefeld, 2003; Tani et al., 2003).

3. Issues motivating the current research and corresponding research questions

The most prominent and overarching issue motivating the current research lies in the need for further empirical evidence to validate possible fundamental processes leading to pro- and antisocial behaviour development. Given this situation, the aim of our research is not to start by evaluating the effectiveness of an intervention program discouraging antisocial behaviour and/or strengthening prosocial behaviour with yet another positive or negative outcome. Instead, the starting point of our research concerns the question what underlies anti- and prosocial behaviour in middle childhood. Specifically, we argued that moral functioning might serve as a central process underlying children's anti- and prosocial behaviour in elementary school. In line with this, the present research pursues a greater understanding of the moral processes underlying antisocial and prosocial behaviour. Since the Four Component Model of Rest's (1983; 1986)

currently offers the most adequate framework to examine the underlying psychological processes of moral behaviour, we rely on this framework. Moreover, research into moral processes rarely focussed on all moral components simultaneously. Up until now, there is scant empirical evidence of the relations between the four described components of moral functioning and their association with moral behaviour (Hardy, 2006; Morton et al., 2006). This gap is unfortunate because it results in fragmented knowledge about the general process of the development of morality. By uncovering the whole scope of moral functioning in pupils age 7-12 this research could guide the development of more effective educational interventions that affect bullying and anti- and prosocial behaviour. The continuing decline in formerly coherent value systems and an increasing individualization in modern Western society make this even more relevant (Rupp & Veugelers, 2003).

In the light of getting more insights into moral functioning as a fundamental process underlying anti- and prosocial behaviour in middle childhood, it is crucial that assessment tools continue to be optimized and tested. Moreover, the operationalization of the Four Component Model is a stringent undertaking (Bebeau, 1994). Therefore, the starting point of our project was the operationalization of the moral components as defined by Rest. Most attention was given to the assessment of moral motivation since small variations in the assessment of these concepts were known to have a strong influence on children's replies (Malti & Ongley, 2014; Nunner-Winkler, 2013). Moreover, the psychometric properties of the assessment of anticipated emotions in the context of moral transgressions, i.e. our operationalization of moral motivation, was not systematically examined yet (Arsenio, 2014). However, supportive evidence of reliability and validity clearly is a critical feature of meaningful research. Our first research question therefore concerns the examination of the validity and reliability of the assessment of emotions in the context of moral dilemmas. We tested this assessment method in the light of two of its important aspects, its domain and developmental variability. Also, we looked at its links to important criterion measures, namely aggressive and prosocial behaviour, and sympathy.

Research Question 1: What is the validity and reliability of the assessment of emotions in the context of moral dilemmas?

Another issue motivating our research lies in the need for further empirical evidence to validate models of possible processes leading to pro- and antisocial behaviour development. Only by examining the way in which development takes place (educational) interventions might be better able to affect both pro- and antisocial behaviour. Moreover, only by better understanding the underlying processes of the development of anti- and prosocial behaviour interventions might be able to steer them into a less antisocial and/or more prosocial career. Remarkably, we found a gap in the research trying to explain the development of anti- and prosocial behaviour. According to Caprara, Dodge, Pastorelli and Zelli, (2007) marginal deviations in behaviour have been neglected in past research, which focused mainly on extreme groups or continuous dimensions of behaviour. The theory of marginal deviations argues that marginal deviations in behaviour, i.e. a small degree of reliable variability in behaviour between a particular child and a normative baseline, also have the potential to develop into higher or lower levels of this behaviour (Caprara et al., 1992; Caprara et al., 2007). Moreover, the theory concerns the processes through which behaviour that is initially only marginally deviant from the norm get transformed into higher or lower levels of this behaviour over time (Caprara, 1992; Caprara & Pastorelli, 1993; Caprara & Zimbardo, 1996). Even though the theory is said to apply to growth in both antisocial and prosocial behaviour, empirical evidence for this proposition is lacking. Indeed, the processes underlying prosocial behaviour might be quite different from processes underlying antisocial behaviour (Baumeister, Bratslavsky, Finkenauer & Vohs, 2001; Krueger, Hicks & Mcgue, 2001; Malti & Krettenauer, 2013; Rothbart & Park, 1986). This motivated us to study marginal deviations in prosocial behaviour, providing new insights in research and theorizing about prosocial development. We were interested in prosocial behaviour development in marginally prosocial children as well as marginally nonprosocial children, respectively initially having a standard score greater than 0 and less than 1 SD above the norm and smaller than 0 and more than 1 SD below the norm of prosocial behaviour.

An interesting avenue for studying the development of marginal deviations in prosocial behaviour into higher or lower levels of prosocial behaviour concerns the effect of (the accumulation of) moral components. Since the separate moral components are understood to be related to prosocial behaviour, they might also contribute to the transformation of marginally deviations in prosocial behaviour into higher or lower levels of prosocial behaviour. Moreover, the theory of marginal

deviations states that the accumulation of personal protective or risk factors might support the development of marginal deviant behaviour into higher levels of this behaviour. For example, a child's marginal deviations in prosocial behaviour could be coupled with high social perspective taking skills and low aggressive behaviour. In this way protective factors accumulate. Moreover, Caprara et al. (2007) showed that a combination of marginal deviations in aggression and accumulated risk factors increase the chance of the development of aggression. In a similar way, the accumulation of moral components might induce the development of prosocial behaviour. This is in line with the Four Component Theory of Rest (1983; 1986) since this theory states that all moral components, i.e. psychological processes, must be in place in order to act moral. However, thus far, it remains unexplored whether and which moral factors moderate the effect of marginal deviations in prosocial behaviour on the development of prosocial behaviour. Also, it is unclear whether the accumulation of central moral processes plays a role in predicting whether marginal deviations in prosocial behaviour develops into higher or lower levels of prosocial behaviour. Our second research question therefore concerns the investigation of the theory of marginal deviations in relation to prosocial behaviour and (the accumulation of) moral components.

> *Research Question 2: What is the relative contribution of (the accumulation of) the four moral components to the development of prosocial behaviour in the light of the theory of marginal deviations?*

A misconception that further triggered our research is that prosocial and antisocial behaviour have been viewed as opposite ends of a single dimension (Hawley, Little & Card, 2007). However, numerous studies have shown that the (longitudinal) correlations between antisocial and prosocial behaviours are at best modest in early to middle childhood (e.g. Carlo, Hausmann, Christiansen & Randall, 2003; Crick & Grotpeter, 1995; Eisenberg et al., 1998; Hastings, ZahnWaxler, Robinson, Usher & Bridges, 2000; Howes & Phillipsen, 1998; Hughes & Dunn, 2000; Tremblay, Vitaro, Gagnon, Piche & Royer, 1992; Wyatt & Carlo, 2002; Zhou et al., 2002). Also, the processes underlying prosocial behaviour might be quite different from processes underlying antisocial behaviour (Baumeister et al., 2001; Krueger et al., 2001; Malti & Krettenauer, 2013; Rothbart & Park, 1986). Therefore, the contribution of the four moral components to pro- and antisocial

behaviour might also not be the same. This motivated us to study the contribution of the moral components to antisocial behaviour, in addition to studying the contribution of moral components to prosocial behaviour with Research Question 2.

In the context of the Four Component Model, bullying seemed a particularly interesting reflection of antisocial behaviour, prevailing among 15% of primary school children (Veenstra et al., 2005). Bullying in school is not only a great burden for teachers and classroom climate, but also for the perpetrators and victims (Hawker & Boulton, 2000; Nansel et al., 2004; Nishina & Juvonen, 2005; Salmivalli & Isaacs, 2005). Since the definition of bullying includes both intentionality and the repetition of a harmful act, bullying has been more strongly related to moral functioning than other forms of antisocial behaviour (Gasser, Gutzwiller-Helfenfinger, Latzko & Malti, 2013). However, research directed at (opposition to) bullying behaviour as an operationalization of moral behaviour in children generally did not explicitly use the Four Component Model as defined by Rest (1983; 1986). Moreover, a major limitation of research relating moral components to children's behaviour in bullying situations has been the neglect of the complex group nature of bullying in schools (Barchia & Bussey, 2011; Faris & Ennet, 2012; Huitsing & Veenstra, 2012; Salmivalli, 2010). The result of this is twofold. First of all, studies have rarely examined the associations between moral components and all the different roles in the bullying process, i.e. the victim, bully, assistant, reinforcer, defender and outsider. Second, researchers have typically focused on individual moral characteristics, whereas moral characteristics of the group have been neglected. This while processes of social influence among classmates can be rather persuasive for the behaviour of children (Espelage, Holt & Henkel, 2003; Juvonen & Galvan, 2008). Research Question 3 therefore advances current research by obtaining insights into the relative contribution of individual and class moral components to bullying as a group phenomenon.

> *Research Question 3: What is the relative contribution of the four components at the individual and class level to bullying-related behaviour?*

Continuing the examination of the underlying moral processes of pro- and antisocial behaviour, we wanted to guide the development of educational interventions that aim at affecting bullying and anti- and prosocial behaviour. Following the outcomes of the previous research questions identifying the relations

between the components of moral functioning and pro- and antisocial behaviour, we plan to distillate the most promising success-promoting moral component for educational interventions. Based on this moral component we will develop a class-based intervention program aimed at the promotion of this specific moral component. In this way, we want to evaluate whether the promotion of this moral component will lead to the reduction of bullying-related and antisocial behaviour and to the promotion of prosocial behaviour in education. Hence, we tried to overcome the current shortcoming of the research into the effectiveness of interventions, i.e. that they focus on what works instead of understanding the process of how an intervention works (Gravemeijer & Kirschner, 2007). Thus, in our fourth research question, we wanted to examine how education, one of the most important contextual factors in childhood, might be able to influence children's functioning on the most promising moral component in order to target bullying-related behaviour as well as anti- and prosocial behaviour in middle childhood. Specifically, we strived to investigate the effects of a class-based intervention program promoting the most prominent moral component on bullying-related behaviour, and anti- and prosocial behaviour over the course of two school years.

> *Research question 4: What are the effects of a class-based intervention program promoting the most promising moral component in middle childhood on bullying-related behaviour, and anti-and prosocial behaviour?*

4. The data and design

In order to answer our research questions data was collected using a four-occasion longitudinal design. The longitudinal study consisted of two parts, as can be seen in Table 1. The first part concerned the first measurement occasion in September/October 2014 and the second measurement occasion in May 2015 with a sample of respectively 1258 and 1261 children in 54 school classes (grade 1 to 6) in 11 primary schools. A total of 1231 children participated in both the first and the second measurement occasion. We took great care in assuring variability in school denomination (three catholic, two protestant, six public), size (from 37 to 307 children, M=119.2 and SD=71.9), location (seven in rural areas and five in cities), and mixing of grades (five with single-graded classroom, six with multi-graded classrooms). During the first part of our study grades 1 to 6 of one school were

offered a class-based intervention promoting agreeableness in between the first and second measurement occasion. Within the same school the parallel grades 1 to 6 were the control condition. The class-based intervention was especially developed and carried out as part of the present research project.

Table 1

Four occasion longitudinal design with two parts

	First part		**Second part**	
	Sept/Oct 2014	May 2015	Sept/Oct 2015	Mar/Apr 2016
Participants				
# schools	11	11	1	1
# classes	54	54	10	10
N	1258	1261	316	312
Intervention	150		161	
Measurements				
Moral sensitivity	x	x	x	x
Moral reasoning	x	x		
Moral motivation	x	x		
Moral character	x	x	x	x
Antisocial behaviour	x	x	x	x
Prosocial behaviour	x	x	x	x
Bullying-related behaviour	x	x	x	x

The second part concerned the third measurement occasion in September/October 2015 and fourth measurement occasion in March/April 2016. Data was collected at the school participating in the class-based intervention in the first part, leading to a subsample of the first two measurement occasions. The third measurement occasion concerned 316 children in 10 school classes from grade 1 to grade 6 in one primary school, among which 227 children that also participated in the first two occasions. The children in grade 6 in the schoolyear 2014 – 2015 left school and new first grade entered the data collection during the school year of 2015 – 2016. Grades 1-5 from 2014 -2015 became grade 2-6 in the school year of 2015 – 2016. Most children participating in the third occasion also participated in the

fourth occasion. There were four children leaving school resulting in a total of 312 children participating in the fourth measurement occasion. During the second part of the study the class-based intervention was again implemented in the same school in between the third and fourth measurement occasion.

At all four measurement occasions, children filled in an online questionnaire assessing aspects of moral sensitivity and moral character, and peer nominations on anti- and prosocial behaviour and bullying-related behaviour. Furthermore, at the first and second occasion all children participated in an one-on-one interview assessing the constructs of moral reasoning and moral motivation. The questionnaires and interviews were administered by trained undergraduate and graduate students. The children were instructed to provide their own responses to the questions and were informed that there were no right or wrong answers. Care was taken to assure children that their answers would remain strictly confidential. Additionally, teachers filled in a short questionnaire about the children in their class concerning aspects of moral character. The study had a very high response rate of varying from 97.3% to 99% for the different measurement occasions. This allowed us to get a complete picture of the whole scope of moral processes and anti- and prosocial behaviours going on within a classroom context over time (Neal, 2008).

5. This dissertation

The present dissertation encompasses four different studies, which were motivated by different, but complementary ideas. Across all four studies we explored the association between moral components and anti- and prosocial behaviour. Before getting to the gist of it, we believed it was important to find adequate measures of the concepts under study. Chapter 2 reflects this notion by presenting a detailed account of the assessment of moral motivation answering Research Question 1. This chapter describes the reliability and validity of the assessment of anticipated emotions in the context of moral transgressions with a special interest in the domain and developmental specificity of the instrument. The instrument consisted of six transgression scenarios covering three domains: unfairness (not winning fairly, not keeping word), omission of a prosocial duty (not sharing, not helping) and victimization (verbal bullying, relational bullying). To evaluate the concurrent and predictive validity, we also examined the relation

between anticipated emotions and antisocial and prosocial behaviour and sympathy at two time points.

In Chapter 3 and 4 we examined the Four Component Model of Rest in two different but complimentary ways. Chapter 3 used the theory of marginal deviations to look into children's development of prosocial behaviour answering Research Question 2. Specifically, we studied whether moral sensitivity, moral reasoning, moral motivation and moral character influence whether marginal deviations in prosocial behaviour accelerates into higher or lower levels of prosocial behaviour. With the help of a multilevel regression analysis we examined possible processes leading to prosocial behaviour development. Further, in Chapter 4, moral behaviour concerned bullying-related behaviour according to five different participant roles: bullying, assisting, defending, victimization or being a bystander. Specifically, we looked into the relative contribution of the four moral components to bullying as a group process, answering Research Question 3. All the different roles in the bullying process were compared with regard to moral sensitivity, moral reasoning, moral motivation, and moral character at both the individual and class level using a multinomial multilevel analysis.

Then, Chapter 5 presents the effects of an intervention aimed at promoting agreeableness in middle childhood answering Research Question 4. Specifically, this paper investigates the effects of a class-based intervention program promoting agreeableness on bullying, assisting, defending, victimization, outsider behaviour, prosocial behaviour, and antisocial behaviour over the course of two school years. Some children received the intervention program during the first year, some during the second year, and some children received the intervention program during both years. The effects of the intervention program were evaluated with the help of multilevel growth curve modelling comparing the different intervention groups to the control group on the growth of the aforementioned outcomes.

We close this dissertation with Chapter 6, in which we recapitulate and discuss the main findings from our four empirical studies, we analyse their limitations, and we derive implications for practice and further research. Although the studies mainly reason about general processes, the conclusions drawn have concrete implications for practitioners and policymakers who want to discourage antisocial behaviour or stimulate prosocial behaviour in the context of primary education. Through our focus on processes, we were able to present some general ingredients for the development and adjustment of intervention programs.

Chapter 2 Assessment of Anticipated Emotions in Moral Transgressions

Abstract

This paper describes the reliability and validity of the assessment of anticipated emotions in the context of moral transgressions in a sample of 1179 children age 6-13 (M=9.1; SD=1.8, 49.0% girls), with a special interest in the domain and developmental specificity of the instrument. To evaluate the concurrent and predictive validity, we also examined the relation between anticipated emotions and antisocial and prosocial tendencies and sympathy at two time points. The instrument consisted of six transgression scenarios covering three domains: unfairness (not winning fairly, not keeping word), omission of a prosocial duty (not sharing, not helping) and victimization (verbal bullying, relational bullying). Results show sufficient internal consistency and a one-factor structure of the anticipated emotions, indicating a lack of domain variability of the assessment of anticipated emotions. Additionally, emotions following hypothetical moral transgressions showed some developmental variability. Whereas no relation was found between anticipated emotions and antisocial tendencies, anticipated negative emotions following the moral transgressions were positively related to prosocial tendencies and sympathy. This provides preliminary evidence for the concurrent and predictive validity of the instrument.

Note. This chapter is based on Jansma, D.J., Malti, T., Opdenakker, M.C.J.L. & Van der Werf, M.P.C. (in press). Assessment of Anticipated Emotions in Moral Transgressions. *European Journal of Psychological Assessment*.

1. Introduction

Children's emotions greatly influence their ability to understand, differentiate, and remember moral and other types of social events. Emotions are short, intense episodes that occur in response to an event or person (Schwarz & Clore, 1996) and are expected to be predictive of children's antisocial and prosocial behaviour (Arsenio, Gold & Adams, 2006; Lemerise & Arsenio, 2000). A central approach to the study of children's emotions in the context of moral and social conflict has been to focus on the anticipation of emotions following moral transgressions, i.e. violations of a moral norm (Malti & Krettenauer, 2013). The underlying assumption is that children's emotion attributions represent authentic expressions of what is important to them in a given moral conflict scenario (Malti, 2016; see Nunner-Winkler, 1999). Interestingly, in spite of the fact that assessment of anticipated emotions following the transgression of moral rules has been in existence since the '80's, little work thoroughly evaluated the reliability and (aspects of) the validity of the assessment of children's self-evaluative emotions (Arsenio, 2014). This is surprising considering the scientific and practical relevance of good assessment in the area of emotions in everyday moral conflict. Up until now, however, several studies did contribute to the improvement of the assessment tool by comparing alternative ways of assessment (for a detailed overview see Malti & Ongley, 2014; Nunner-Winkler, 2013). For example, Keller, Lourenço, Malti and Saalbach (2003) showed that children more often attribute negative emotions to the self than to hypothetical wrongdoers. Yet, in the light of creating and evaluating interventions targeted at behaviour in social and moral events, it is crucial that assessment tools continue to be optimized and tested. Moreover, supportive evidence of reliability and validity is clearly a critical feature of meaningful research. Therefore, the current study examines the validity and reliability of the assessment of emotions in the context of moral transgressions in the light of two important aspects of this particular assessment method, i.e. domain and developmental variability, and its links to important criterion measures, i.e. aggressive and prosocial tendencies, and sympathy.

1.1 Assessing anticipated emotions following moral transgressions

In the recent paradigm emotions following moral transgressions are assessed in a one-on-one interview procedure in which children are provided with

stories concerning hypothetical moral and social transgressions, such as inflicting harm on others, omitting prosocial duties, excluding others, etc. The interview presents several moral transgression scenarios in which the main character in the story violates a social or moral norm in order to achieve a personal goal. A moral norm "bears on the interest or welfare either of society as a whole or at least of persons other than the judge or agent" (Gewirth, 1984, p. 978). Following each scenario, children are asked how they would feel if they would be the one who transgressed the moral rule or obligation. Their spontaneously mentioned emotions are generally divided into negative, positive and mixed emotions and are indicative for the importance children attach to moral norms in contrast with need satisfaction. By attributing an emotion children indicate which of the two facts that are simultaneously true of a wrongdoer – that s/he transgressed a moral norm and satisfied a personal desire – they deem more important (Nunner-Winkler & Sodian, 1988). Whereas negative emotions, such as sadness, and shame, are said to indicate guilt feelings and are seen as indicators of moral awareness, positive emotions, such as happiness, pride, and joy, are seen as indicators of a lack of moral awareness (Malti et al., 2009; Malti & Ongley, 2014). Mixed emotions, a little good and a little bad, point to the understanding that a moral transgression may cause positive and negative emotions at the same time (Mascolo & Fischer, 2010). Therefore, mixed emotions also indicate moral awareness (Keller et al., 2010; Perren, Gutzwiller-Helfenfinger, Malti & Hymel, 2012).

1.2 Domain variability in children's anticipated emotions in moral transgressions

Not surprisingly, children's anticipated emotions following a hypothetical moral transgression strongly depend on the specific scenario presented to them (e.g. Arsenio, 1988; Arsenio & Fleiss, 1996; Smetana, Compione-Barr & Yell, 2003; Wiersma & Laupa, 2000). Therefore, it is crucial to pick a variety of scenarios. An important selection criterion concerns the domain the scenario is in. Domains that are often distinguished are the neglect of a prosocial duty (e.g. not sharing), unfairness (e.g. not keeping word) and victimization (e.g. harming others) (Arsenio, 2014). Most researchers assess children's anticipated emotions using a mix of stimulus stories involving both acts of victimization, unfairness, and refusing to act prosocial (Arsenio, 2014) and collapse data for these different types of transgressions (Malti & Krettenauer, 2013). However, children generally see neglect of a positive prosocial duty as less objectionable and creating less psychological

harm than actually harming others or being unfair (Malti, Gasser & Buchman, 2009; Miller, 2006; Smetana, 2006; Turiel, 2006). In line with this, Arsenio (2014) argues that researchers should separate findings for emotions following harm and unfair treatment and emotions following a prosocial violation. The present study therefore examines whether the different domains of moral violation should be structurally distinguished.

1.3 Developmental variability in children's anticipated emotions in moral transgressions

Children's anticipated emotions in moral transgressions strongly depend on development. Already in the earliest studies (Arsenio, 1988; Arsenio & Ford, 1985; Barden, Zelko, Duncan & Masters, 1980) striking age-related differences for emotions following moral transgressions were found. It is well established that the occurrence of negative emotions following a hypothetical transgression show an increase after age 6-7 years (Arsenio et al., 2006; Krettenauer, Malti & Sokol, 2008). The percentage of positive emotions ranges from 93% (Arsenio & Lover, 1995) to 55% (Nunner-Winkler, 2008) among 6-7 year olds, and from 67% (Lourenco, 1997) to 35% (Nunner-Winkler, 2008) among 8-9 year olds. The large discrepancies in positive emotions attributed in the same age group are likely a result of variations in assessment procedures (Nunner-Winkler, 2013). Generally, the percentages show a pattern of an increased attribution of negative emotions. However, for some individuals, the attribution of positive emotions remains well into adolescence, pointing to stable inter-individual differences in the anticipation of emotions later in life (Arsenio et al., 2004; Krettenauer & Eichler, 2006; Malti & Ongley, 2014). The current study will take the apparent developmental variability of children's anticipated emotions in moral transgressions into account when examining the structure and validity of the assessment of these emotions.

1.4 Concurrent and predictive validity of anticipated emotions in moral transgressions

The predictive validity for emotions in moral transgressions can be checked quite nicely by relating them to behavioural tendencies. Studies of children's anticipated emotions generally assume that children remember the emotional antecedents and consequences of social situations and that this connection between events and emotions guides future behaviour (Arsenio, 2014). Negative emotions

following moral transgression are said to provide an early foundation for the development of other-oriented behavioural tendencies, because they indicate that the self feels committed to a moral standard (Malti, Gummerum, Keller & Buchmann, 2009; Malti & Krettenauer, 2013; Tangney et al., 2007). Specifically, anticipated negative emotions restrict aggressive behaviour and motivate reparative behaviour, such as apologies (Asendorpf & Nunner-Winkler, 1992; Malti & Keller, 2010; Tangney et al., 2007). In a recent meta-analysis, Malti and Krettenauer (2013) showed positive small-size relations between negative emotion attributions and prosocial behaviour and negative moderate-size relations between negative emotion attributions and antisocial behaviour independent of age. Next to showing the association between emotions and behaviour, these results indicate that children's anticipated emotions are more strongly related to antisocial than to prosocial behaviour. This might be due to the fact that hypothetical scenarios in research on anticipated emotions present rule transgressions and actions that are harmful to others. Therefore, assessment of anticipated emotions in moral transgressions reflects emotions primarily in domains of antisocial behaviour. Interestingly, the way emotions are assessed may influence their relation with social behaviour. That is why the current study will relate pro- and antisocial tendencies of children to their anticipated emotions in transgressions in different domains.

Additionally, the concurrent and predictive validity of the assessment of anticipated emotions following moral transgressions will also be tested using a measure of sympathy. Links between sympathy and anticipated emotions following moral transgressions are well established (Malti et al., 2009). Sympathy is "a vicarious emotional reaction that is based on the apprehension of another's emotional state or situation, which involves feelings of sorrow or concern for the other" (Eisenberg, Shea, Carlo & Knight, 2014, p.65). Both sympathy and negative emotional reactions to moral transgressions arise from the concern for the other's distress (Tangney & Dearing, 2002).

1.5 The present study

The aim of the current study is to look into the domain and developmental variability of the assessment of children's anticipated emotions in moral transgressions and their links to aggressive and prosocial tendencies and sympathy. To this end, we used a mix of scenarios involving acts of unfairness, victimization, and omission of prosocial duty, to determine if these together could be considered

as a valid and reliable measure of these emotions. The present study evaluated the proposed instrument with respect to its reliability and empirical structure using categorical principal component analysis. Domain and developmental variability were taken into account. Second, the concurrent and predictive validity of the measure was evaluated. We examined the links of anticipated emotions to aggressive and prosocial tendencies at two time points. As an additional check of the concurrent and predictive validity of the measure of anticipated emotions, self-reported sympathy was related to anticipated emotions at two time points. The role of both domain and developmental variation in responses was again taken into consideration.

With regard to the measurement model we were particularly interested to see whether a one- or three-factor structure of the instrument would best fit the data. The custom of collapsing the reported emotions from different scenarios suggests that the different scenarios have only one underlying dimension indicating that the self feels committed to moral standards. However, taking into account the claim of Arsenio (2014) that children's anticipated emotions depend on the specific domain of the transgression suggests that the structure might depend on the domains used in the instrument. In our case, three different domains were used, namely unfairness and victimization, and omission of prosocial duty. Thus, a three-factor structure would provide evidence for the domain variability of the assessment of anticipated emotions. Notwithstanding the lack of reports of reliability estimates, the custom of collapsing the data from different scenarios also suggests that the different scenarios provide a consistent measurement of individuals. Thus, we further hypothesized the reliability of the six scenarios to be satisfactory. With regard to the developmental variability, the anticipation of negative and mixed emotions was expected to increase after age 6-7 years. However, we expected to find the same factor structure and reliability across age. Based on the meta-analysis of Malti and Krettenauer (2013), children's anticipated negative and mixed emotions were further expected to be positively related to prosocial tendencies and negatively to antisocial tendencies across different age groups. Depending on the outcome of the factor analysis, anticipated emotions in the domain of victimization and unfairness were expected to relate more (negatively) to aggressive than (positively) to prosocial tendencies since these scenarios reflect the domain of rule transgressions and actions that are harmful to others. On the other hand, anticipated emotions in the domain of the omission of

prosocial duties were expected to relate more to prosocial than to aggressive tendencies. A positive relation was expected between sympathy and children's negative and mixed emotions following hypothetical transgressions (Malti et al., 2009). Since sympathy is related to prosocial tendencies rather than antisocial tendencies (Eisenberg et al., 2006), sympathy was thought to relate more to anticipated emotions in the domain of the omission of prosocial duties than to anticipated emotions in the domain of victimization and unfairness in case more than one factor was identified. No age differences were expected.

2. Method

2.1 Participants

The sample used in this study originates from a longitudinal research project that was designed to examine moral functioning and prosocial and antisocial behaviour of school-aged children. Table 1 shows the sociodemographic and clinical data for the sample. Ethical consent for the longitudinal research project was obtained from the Ethical Committee Pedagogical and Educational Sciences from the University of Groningen. Participants were recruited via the personal network of the researchers. First, school principals and teachers were asked for consent. Parental consent letters were than distributed to obtain permission for their

Table 1

Sociodemographic and clinical data for the sample (N=1179)

Age	*M*=9.1; *SD*=1.8
Sex	
- Girls	49.0%
Ethnicity	
- Dutch	94.3%
Education level of parents*	
- Low	23.4%
- Middle	47.7%
- High	29.9%
Diagnosed disorders	
- All	7.7%
- ADHD	1.7%

* International Standard Classification of Education (UNESCO, 2011)

children's participation (acceptance rate: 99%). The 1179 (601 boys and 578 girls) children attended 11 different elementary schools in the North of the Netherlands divided over 52 classrooms. The ages ranged from six up to and including 13 years (M=9.1; SD=1.8). The children attended regular education and comprised of predominantly white pupils of Dutch descent.

2.2 Materials and Procedure

All children filled in an online questionnaire and participated in an one-on-one interview. All questions and measures were formulated and posed in Dutch. Both the online questionnaire and the interview were administered in the autumn of 2014, and the summer of 2015 by graduates from educational and pedagogical sciences together with Bachelor students from the same program. The graduates and students all received a training of two days. The children were instructed to provide their own responses to the questions and were informed that there were no right or wrong answers. Great care was taken to assure students that their answers would remain strictly confidential and would not be revealed to anyone else and accordingly, each child was given an anonymous ID number generated automatically in the coded data set.

Anticipated emotions following moral transgressions. In the autumn of 2014 children's anticipated emotions following moral transgressions were assessed. To this end, we modified six scenarios depicting hypothetical transgressions covering three moral domains: unfairness (not winning fairly, not keeping word), victimization (verbal bullying, relational bullying) and omission of prosocial duties (refusing to share pencils, refusing to help someone in pain). The modification of the scenarios involved both the translation in Dutch as well as adapting the scenarios to age groups and cultural context. The selection of the six stories was based on three studies, N=121, N=88 and N=145, piloting and validating fourteen scenarios. These fourteen scenarios were based on previous studies (i.e. Davidson, Turiel & Black, 1983; Keller et al., 2003; Krettenauer & Eichler, 2006; Malti, Gummerum, Keller & Buchmann, 2009; Malti & Keller, 2010; Malti, Ongley, Dys & Colasante, 2012; Nunner-Winkler, 1999; Nunner-Winkler & Sodian, 1988; Nucci, 1981; Olthof, Schouten, Kuiper, Stegge & Jennekens-Schinkel, 2000; Smetana, 1981) and were selected because of their variety in content, domain and severity of the transgression, and because there were indications of their validity. The stories were gender-matched and illustrated with cartoons. Participants could not judge the

Table 2

The domain and content of the scenarios used for the assessment of emotions in moral transgressions

Domain of scenario	Content of scenario	Content of conflicting issue
Unfairness	Not winning fairly	Having more candy
Unfairness	Not keeping word	Earning more money
Omission of prosocial duty	Not sharing pencils	Making a drawing for a friend
Omission of prosocial duty	Not helping someone in pain	Attending the birthday party of a friend
Victimization	Verbal bullying	Belonging to the popular group
Victimization	Relational bullying	Becoming victimized

characters' emotions by looking at the pictures, since their face was not visible after the transgression. No adults were involved to avoid confounding children's moral understanding with their understanding of authority. Following each scenario, children were asked how they would feel if they would have been the person who transgressed the moral rule (anticipated emotion) and why they would feel that way (justification for anticipated emotion). The interview took 15-25 minutes. The child sat at the side of the test administrator with the cartoons in front, so they could see them clearly. The cartoons remained in front of the child during the interview in order to prevent mistakes due to memory requirements. The interviews were recorded and transcribed afterwards. Table 2 gives an overview of the scenarios and the conflicting issues involved. In the sharing story, for example, the moral issue concerned sharing whereas the conflicting issue concerned making a drawing for a friend. Girls were read the following text: 'This is Samina. Samina is doing crafts in the hall of the school. She found all the good school pencils before getting started. She wants to craft something nice for her friend and that is why she needed the good pencils. Fenna, a girl from another class, comes up to Samina and says: "Hi Samina, I would also like to craft. May I take half of your pencils?" Samina thinks for a moment and replies: "No, you may not have the pencils. I would like to use all them myself." Samina continues crafting. Fenna walks away sad.' After the story they were asked the questions 'How would you feel if you were Samina?' and 'Why would you feel […] if you were Samina?'. When the answer to the first question was ambiguous, children would be asked whether they

would feel more positive or more negative. Matching cartoons for the sharing scenario are provided in Appendix A.

Following Perren, Gutzwiller-Helfenfinger, Malti and Hymel (2012), anticipated emotions were coded as negative (e.g., bad), mixed (e.g. half well and half bad) or positive (e.g., happy) emotions. First, the spontaneously mentioned emotions were provided with category labels of nine emotions derived from the social-emotional responding task (SERT) coding scheme (Malti, 2016): happy, neutral, angry, scared, bad, sad, guilty, ashamed, or mean/disgusted. The exact coding scheme can be found in Appendix B. Then, *angry, scared, bad, sad, guilty, ashamed,* and *mean/disgusted* responses were coded as negative emotions, whereas *happy* responses were coded as positive emotions. In line with other research, *neutral* responses (ranging from 0.2% to 1.6%) were left out of the analysis because we had no specific hypotheses regarding these types of emotions (Malti, Killen & Gasser, 2012). Mixed emotions were coded if children mentioned both negative and positive emotions. Intercoder reliability of the coding of emotions was calculated with Cohen's Kappa, resulting in K=.97 for not winning fairly, K= .94 for not keeping word, K= .97 for not sharing pencils, K= .99 for not helping someone in pain, K= .94 for verbal bullying and K= 1.00 for relational bullying. The raters discussed disagreement with each other until a consensus was reached and the consensus was then coded.

Behavioural measures. An online peer nominations questionnaire was used to collect data on aggressive prosocial tendencies in the autumn of 2014 and in the summer of 2015. The online questionnaire was preceded by a general introduction and instructions regarding the measure. The children were seated in groups of four to ten at computers spaced sufficiently to ensure privacy. At all times, a trained test administrator was available for help. Students in grade 1 and those with difficulty concentrating and/or reading were assessed one-on-one, where the researcher read out the questionnaire. In the questionnaire, peer ratings of aggressive ('Who quarrels a lot?' and 'Who says and does mean things?') and prosocial tendencies ('Who helps other children?' and 'Who says and does nice things?') were obtained. The children were given a roster with the names of their classmates and they could select as many, or as few, classmates as they wanted. This sociometric method of assessing behaviour has been used in other studies that established its validity (prosocial behaviour: Carlo, Koller, Eisenberg, Da Silva & Frohlich, 1996; Deković & Janssens, 1992; antisocial behaviour: Newcomb,

Bukowski & Pattee, 1993; Veenstra et al., 2005). The frequency with which each participant was nominated was divided by the number of classmates who were nominating, in order to adjust for class size. Cronbach's alpha for aggressive and prosocial tendencies was .90 and .75 for the first measurement occasion and .89 and .77 for the second measurement occasion, respectively.

Sympathy. In the autumn of 2014 and in the summer of 2015 children's self-reported sympathy was measured with 9 statements based on the validated empathic concern measure of Eisenberg, Fabes, Murphy, Karbon, Smith, and Maszk (1996). After each statement (e.g. "When I see someone being picked on, I feel sorry for them") children were asked whether the sentence described him/her or not, and if so, how strongly: 1="*No, this does not sound like me*", 2= "*This is sort of like me*", and 3= "*This is really like me*". Cronbach's alpha was .81 for the first measurement occasion and .84 for the second measurement occasion.

2.3 Statistical Analysis

First, the percentages of anticipated negative, mixed and positive emotions following the six scenarios were examined per domain and developmental period to get a feeling for the developmental and domain variability if anticipated emotions in moral transgressions. After that, the structure of the anticipated emotions following moral transgressions was analyzed. In order to provide preliminary validity support for the instrument, a two-level categorical exploratory factor analysis was executed using Mplus Version 7.31 (Muthén & Muthén, 2015). A two-level analysis consists of lower-level observations (i.e. level 1 units of analysis) nested within higher-level observations (i.e. level 2 units of analysis). In the context of the present study, pupils were nested within classrooms. The design effect, indicating how much the deviation from independent observations will influence the standard errors in the analysis, was 2.67. A design effect greater than 2 indicates that the clustering in the data needs to be taken into account during estimation using multilevel analysis (Snijders & Bosker, 2011). Additionally, confirmatory factor analysis is a popular approach to come up with a small set of underlying dimensions of an instrument (i.e. factors or components) that represent most of the information found in the original items (Fabrigar, Wegener, MacCallum & Strahan, 1999). Considering our interest in domain variability, a comparison had to be made between a one factor structure and a three factor structure since our instrument included three different domains. However, confirmatory factor models could not

be identified because the three factor model includes only two items per factor. A minimum of three items must load significantly on each factor for all of the subscales to be successfully identified (Little, Lindenberger & Nesselroade, 1999; Velicer & Fava, 1998). Therefore, an exploratory factor analysis for categorical factor indicators using WLSM estimation was more in place. Another issue involved the coding of mixed emotions; mixed emotions do not occur often and were mostly discarded in previous research (Lourenço, 1997; Menesini et al., 2003; Keller et al., 2003). Thus, on top of the theoretical notion that mixed and negative emotions indicate more moral awareness than positive emotions, it was difficult to determine whether mixed emotions would be less, equal to, or more indicative of moral awareness than negative emotions. In the exploratory factor analysis for categorical factor indicators we therefore tested three different options at the within level of the model: coding mixed emotions as less than negative emotions, more than negative emotions and equal to negative emotions. As Gorsuch (1983) recommended, we used multiple methods to determine the best coding system and the number of factors to retain, including the Eigenvalue >1 rule (Kaiser, 1960), commonly used indicators of model fit like the comparative fit index (CFI; Bentler, 1990), Tucker–Lewis Index (TLI; Tucker & Lewis, 1973), and root mean squared error of approximation (RMSEA; Steiger, 1989) (Cheung & Rensvold, 2002), and theory. The default rotation, the oblique rotation of Geomin, was used. Internal consistency of the scale(s) was measured through Cronbach's alpha (Nunnally & Bernstein, 1994). Finally, the results were examined across sex and age to check developmental variability.

Then, the concurrent and predictive validity of the measure was examined taking into account developmental variability. Domain variability was only taken into account in case the exploratory factor analysis would indicate more than one factor. Based on the factor structure, anticipated emotions were related to aggressive and prosocial tendencies and sympathy at two time points using multilevel techniques with pupils nested in classes in MLWiN 2.35 (Rasbash, Browne, Goldstein & Yang, 2000). In all multilevel analyses, age was included as a moderator between anticipated emotion scores and pro- and antisocial tendencies and sympathy to test the developmental variability of this relation. Sex, scholastic ability, and parental education were included as control variables.

3. Results

3.1 Missing data

In order to compute the factor analysis in Mplus and the multilevel analysis in MLwiN, missing data was deleted in a listwise manner. The percentage missing varied from 0.0% to 1.4% for the different scenarios and resulted in a total of 3.7% children missing one or more items due to item nonresponse. The children who answered all questions did not differ from the children with item(s) missing in terms of education level of parents, scholastic ability, antisocial tendencies, prosocial tendencies, sympathy, and emotions following moral transgressions ($t(1151)=1.29$; $p=.20$; $t(1147)=1.29$; $p=.20$; $t(1177)=.-1;71$ $p=.08$; $t(1177)=.10$; $p=.92$; $t(1175)=-.05$; $p=.96$; $t(1177)=.04$; $p=.99$). However, children with items missing were younger ($t(1177)=-.43$, $p<.01$) and comprised a higher percentage of girls ($\chi^2(1)=5.21$; $p=.01$) than the children with complete data.

3.2 Descriptive statistics of the assessment of emotions in moral transgressions

Before conducting the factor analysis and multilevel analysis, the domain and developmental variability of emotions following moral transgressions was examined descriptively. Table 3 presents the percentage of anticipated negative, mixed and positive emotions following the six scenarios per domain and developmental period. Age was split into three developmental periods (6-8 years, 9-10 years, and 11-12 years). The percentage of negative emotions ranges from 64.7 to 85.5, indicating that the vast majority of children attribute negative emotions to the self as hypothetical transgressor. The mean percentage of negative emotions is highest within the domain of the omission of prosocial duties (scenarios: not sharing, not helping) and lowest within the domain of unfairness (scenarios: not winning fairly, not keeping word). However, this difference was not significant ($F(2)=6.65$; $p=0.08$). Mixed emotions did not occur often, 3.6% to 17.2 %, and their mean percentage did not differ significantly across domains ($F(2)=2.49$; $p=0.23$). Positive emotions were least frequent for the scenarios in the domain omission of prosocial duty in comparison to the other domains, 13.2% versus 22.9% ($F(2)=20.77$; $p=0.02$). With regard to developmental variability, children age 6-8 years reported most positive emotions ($F(2)=6.85$; $p<0.01$), whereas children age 9-12 years reported most mixed emotions ($F(2)=22.31$; $p<0.01$) across scenarios.

Table 3

Percentages of anticipated negative, mixed and positive emotions following six scenarios per domain and per developmental period

Domain	Scenario	Developmental period	% negative emotions	% mixed emotions	% positive emotions
Omission of prosocial duty	Not sharing		85.5	3.6	10.9
		6-8 years	83.6	2.1	14.2
		9-10 years	87.8	4.1	8.0
		11-12 years	85.1	4.5	10.4
	Not helping		74.8	9.6	15.5
		6-8 years	71.5	5.3	23.1
		9-10 years	75.9	10.7	13.4
		11-12 years	76.9	12.5	10.6
Unfairness	Not winning fairly		64.7	10.0	25.3
		6-8 years	65.3	7.2	27.5
		9-10 years	65.1	11.3	23.6
		11-12 years	63.9	11.3	24.8
	Not keeping word		56.8	17.2	26.0
		6-8 years	61.6	8.7	29.6
		9-10 years	55.7	18.5	25.8
		11-12 years	53.5	23.6	22.9
Victimization	Relational bullying		75.5	4.0	20.5
		6-8 years	66.9	3.5	29.5
		9-10 years	78.3	4.2	17.5
		11-12 years	80.6	4.3	15.1
	Verbal bullying		73.7	6.7	19.6
		6-8 years	78.1	2.4	19.5
		9-10 years	75.2	6.5	18.3
		11-12 years	68.4	10.6	21.0
	Total		70.8	8.4	19.4
		6-8 years	70.0	4.8	23.5
		9-10 years	72.0	9.1	17.6
		11-12 years	70.8	11.0	17.3

Table 4

Fit statistics for models with one and two between factors, and one and two within factors for coding mixed emotions as less than negative emotions, more than negative emotions and equal to negative emotions

Model	# of free parameters	RMSEA [90% CI]	CFI	TLI
One between factor:	30	.042 [.030; .055]	.950	.916
1a. One within factor – positive is coded as 1, negative as 2, mixed as 3	30	.028 [.014; .042]	.990	.984
1b. One within factor – positive is coded as 1, negative as 3, mixed as 2				
1c. One within factor – positive is coded as 1, negative as 2, mixed as 2	24	.021 [.000; .036]	.993	.988
2a. Two within factors – positive is coded as 1, negative as 2, mixed as 3	35	.036 [.021; .052]	.973	.938
2b. Two within factors – positive is coded as 1, negative as 3, mixed as 2	35	.014 [.000; .034]	.998	.996
2c. Two within factors – positive is coded as 1, negative as 2, mixed as 2	29	.008 [.000; .031]	.999	.998
Two between factors:				
1a. One within factor – positive is coded as 1, negative as 2, mixed as 3	35	.048 [.034; .063]	.953	.890
1b. One within factor – positive is coded as 1, negative as 3, mixed as 2.	35	.028 [.010; .045]	.993	.988
1c. One within factor – positive is coded as 1, negative as 2, mixed as 2	29	.023 [.000; .040]	.994	.986
2a. Two within factors – positive is coded as 1, negative as 2, mixed as 3	40	.045 [.027; .064]	.975	.906
2b. Two within factors – positive is coded as 1, negative as 3, mixed as 2	40	.000 [.000; .028]	1.000	1.004
2c. Two within factors – positive is coded as 1, negative as 2, mixed as 2	34	.000 [.000; .034]	1.000	1.000

3.3 Factorial structure and reliability of the assessment of emotions in moral transgressions

To investigate the factorial structure of the assessment of emotions in moral contexts, multiple exploratory factor analyses for categorical factor indicators were executed. Specifically, we compared three different coding possibilities: coding mixed emotions as less than negative emotions, more than negative emotions and equal to negative emotions. Table 4 reports the number of free parameters, the RMSEA, the CFI, and the TLI for the different models. Throughout the different factor solutions, the best model fit was found when mixed emotions were coded equal to negative emotions. The one within factor one between factor model with mixed emotions coded equal to negative emotions had the least parameters and a relatively good fit. The best fit statistics were found for the two within one between factor combination and the two within two between factor combination with mixed emotions coded as less than or equal to negative emotions. However, the two factor models barely added extra fit. The most parsimonious model is the one within and one between factor solution with mixed emotions coded equal to negative emotions. Table 5 presents the Geomin rotated loadings belonging to this model. All factor loadings were similar and greater than

Table 5

Factor loadings based on an exploratory factor analysis with Geomin rotation for 6 scenarios assessing emotions following moral transgressions separated for sex and age

Scenario	Factor loadings	Factor loadings girls	Factor loadings boys	Factor loadings age 6-8	Factor loadings age 9-10	Factor loadings age 11-12
Not sharing	.76*	.78*	.74*	.70*	.94*	.79*
Not helping	.74*	.81*	.70*	.82*	.65*	.64*
Not winning fairly	.79*	.80*	.77*	.89*	.77*	.74*
Not keeping word	.78*	.81*	.78*	.81*	.78*	.65*
Relational bullying	.76*	.82*	.71*	.67*	.80*	.81*
Verbal bullying	.72*	.75*	.68*	.80*	.81*	.66*
Cronbach's alpha	.76	.79	.71	.80	.75	.70

* Significant at the .05 level.

.70 indicating that all scenarios contribute more or less equally to the content or meaning of the factor. The internal consistency of the scores based on the factor, as measured by Cronbach's alpha, was adequate, namely .76 (Henson, 2001). The factor structure and Cronbach's alpha seem similar for boys and girls, with slightly higher factor loadings and reliability for girls (RMSEA= .000 [.000; .011], CFI= 1.000, TLI=1.000) than for boys (RMSEA= .027 [.000; .048], CFI =.985, TLI =.975). It therefore seems there is no domain variability in the assessment of anticipated emotions in moral transgressions.

With regard to developmental variability, the factor structure and reliability of the six scenarios assessing emotions following moral transgressions was examined for the three developmental periods separately. The RMSEA, CFI, and TLI were .040 [.005; .066], .980, and .967 for the model including children age 6-8; .000 [.000; .023], 1.000 and 1.000 for the model including children age 9-10 and .000 [0.000; .014], 1.000 and 1.000 for the model including children age 11-12. Table 5 shows that the factor loadings and Cronbach's alpha differ across the age groups, but not greatly. Age 6-8 reveals the highest factor loadings and reliability. As we found a one-factor structure that holds across sex and age, we added the scores of the anticipated emotions across the six scenarios and took the mean as an indicator of children's self-evaluated emotions in moral conflict.

3.4 Concurrent and predictive validity of the assessment of emotions in moral transgressions

Table 6 reports the results of the multilevel analysis predicting the mean score of prosocial and antisocial tendencies and sympathy for both time points from anticipated emotions across the moral transgression scenarios while controlling for age, gender, scholastic ability and parental education. Descriptive statistics for the aforementioned relations can be found in Appendix C. Anticipated emotions of children across scenarios were not significantly related to antisocial tendencies at both time points and to prosocial tendencies at time point two. A significant positive relation was found between anticipated emotions and prosocial tendencies at the first time point (β=6.43; $p<0.05$). The association between anticipated emotions and sympathy was significant and positive for both measurement occasions (β=.32; $p<0.01$; β=.20; $p<0.01$).

Table 6

Multilevel regression models predicting antisocial and prosocial tendencies and sympathy from anticipated emotions following moral transgressions

	T1 Antisocial tendencies		T2 Antisocial tendencies		T1 Prosocial tendencies		T2 Prosocial tendencies		T1 Sympathy		T2 Sympathy	
Fixed effects	*Estimate*	*S.E.*	*Estimate*	*S.E.*	*Estimate*	*S.E.*	*Estimate*	*S.E.*	*Estimate*	*S.E.*	*Estimate*	*S.E.*
Intercept	27.64	1.24	26.73	1.26	36.07	1.94	35.56	2.16	1.97	.04	1.92	.04
Level 1 (pupil) variables												
Sex (girl)	-16.85*	1.10	-14.08*	1.09	13.70*	.78	13.90*	.93	.21*	.02	.24*	.03
Socioeconomic status	-3.24*	.69	-2.02*	.68	2.15*	.53	1.91*	.63	.00	.02	.03	.02
Scholastic ability	-3.04*	.52	-3.05*	.52	4.45*	.38	4.50*	.46	.05*	.01	.04*	.01
Age (9-10)	-1.36	1.52	-5.02*	1.54	5.54*	1.63	7.15*	1.92	-.02	.04	-.04	.05
Age (11-12)	-3.11*	1.53	-5.35*	1.57	8.74*	2.10	8.79*	2.44	.06	.05	.03	.05
Anticipated emotions	-6.11	3.47	-2.11	3.46	6.43*	2.52	4.40	2.99	.32*	.09	.20*	.09
Anticipated emotions *Age (9-10)	11.76*	5.19	7.27	5.16	-6.85	3.73	-5.79	4.44	.02	.12	.15	.14
Anticipated emotions* Age (11-12)	5.36	5.09	3.55	5.06	-1.33	3.65	2.33	4.35	-.02	.13	.06	.13
Random effects	*Var.*	*S.E.*	*Var.*	*S.E.*	*Var.*	*S.E.*	*Var.*	*S.E.*	*Var.*	*S.E.*	*Var.*	*S.E.*
Class level variance	5.78	4.18	7.22	4.43	112.85	23.50	130.65	27.65	.01	.01	.02	.01
Pupil level variance	331.55	14.31	325.28	14.07	162.43	7.02	230.01	9.96	.20	.01	.22	.01

* $p<.05$

The above findings were true independent of age, except for the relation between reported anticipated emotions and antisocial tendencies at the first time point. In comparison to the anticipated emotions children age 6-8 report, the anticipated emotions of children age 9-10 show a stronger and positive relation with antisocial tendencies (β=11.76; $p<0.05$). This result directs to developmental variability in the relations between anticipated emotions and the criterion measures.

4. Discussion

This paper presents an instrument for the assessment of children's anticipated emotions following moral transgressions. The instrument consisted of six scenarios covering three domains: unfairness (not winning fairly, not keeping word), omission of a prosocial duty (not sharing, not helping) and victimization (verbal bullying, relational bullying). The reliability and validity of this instrument was examined with a special interest in its domain and developmental specificity. To further evaluate the concurrent and predictive validity of the instrument, we examined the relations between anticipated emotions and aggressive and prosocial tendencies and sympathy at two time points. Again, domain and developmental variability was taken into consideration.

4.1 Structure and reliability of the assessment of emotions in moral transgressions

First of all, the present study offers preliminary evidence for a one-factor structure of the instrument used to assess anticipated emotions in moral transgressions. The most parsimonious model was the model with one within and one between factor solution with mixed emotions coded equal to negative emotions. Based on this result, we decided to dichotomize the anticipated emotions of the children into positive versus mixed/negative emotions. However, the higher fit indices obtained for this model are partly due to the relatively low occurrence of mixed emotions. Interestingly, children who reported mixed emotions (in comparison to positive and negative emotions) show the highest mean scores on prosocial and sympathy and lowest mean scores on antisocial tendencies. Based on these results - that are not reported in the current paper, but available on request - one could also decide to code mixed emotions as higher than negative emotions. Nevertheless, the reported factor analysis and the low occurrence of mixed

emotions plead for the integration of mixed and negative emotions.

There seems to be no domain variability in the assessment of anticipated emotions in moral transgressions. Considering that all scenarios contribute more or less equally to one factor across sex and age, it seems that the different scenarios have one underlying dimension indicating that the child's self feels committed to moral standards. This is consistent with the current custom of collapsing the reported emotions from different scenarios (Malti & Krettenauer, 2013). As expected, the instrument also appeared to be reliable. However, as factor analysis is generally sample dependent, one should be cautious in applying the current findings to other samples with different characteristics. Moreover, one could opt for a more nuanced view on emotions using different scenarios separately. Considering this is the first attempt to look at the reliability and validity of the assessment of emotions in moral transgressions, we encourage future studies to report about the structure and reliability of the instrument assessing anticipated emotions to confirm and complement our findings. Additionally, the small indications for domain variability in the assessment of anticipated emotions - differences in the associations between anticipated emotions and the criterion measures for the different scenarios; differences in the percentages of negative emotions between scenarios - are worth studying further. For example, the mean percentage of negative emotions is highest within the domain of the omission of prosocial duties and lowest within the domain of unfairness. This contrasts with the research of Smetana (2006), Turiel (2006), Miller (2006) and Malti, Gasser and Buchman (2009), indicating that neglect of a positive prosocial duty is generally seen as less objectionable and creating less psychological harm than actually harming others or treating others unfairly. A possible explanation for this adverse finding might be related to the coding of emotions in this study. A more detailed coding scheme, as well as a measure of the intensity of emotions, might stronger reflect the difference in severity between the different domains of moral transgressions (Malti & Krettenauer, 2013). Another explanation might lie in the specific scenarios used in this study. It could be that the scenarios were not contrasting the domains well enough. For example, the scenario about relational bullying clearly involves victimization. However, it also involves the omission of a prosocial duty. Generally speaking, most scenarios used in research on anticipated emotions seem to simultaneously reflect different domains (Nunner-Winkler, 2013).

Developmental variability was found for the structure and reliability of the assessment of anticipated emotions. As expected, children age 6-8 years reported most positive emotions, whereas children age 9-12 years reported most mixed and negative emotions across scenarios (e.g. Arsenio et al., 2006; Krettenauer, Malti & Sokol, 2008). Moreover, the one-factor structure showed the best fit and highest reliability for children age 6-8 years in comparison to the other age groups. However, these differences were relatively small. As Arsenio, Gold and Adams (2004), Krettenauer and Eichler (2006) and Malti and Ongley (2014) suggest, the absence of a strong developmental pattern might indicate stable inter-individual differences in the anticipation of emotions. Since most children in our sample already passed the happy victimizer phenomenon that takes place around age 7-8 years, the developmental pattern might not be apparent in older children. Future studies could try to magnify the conflicting issues in the scenarios in order to increase the number of anticipated positive emotions and avoid a ceiling effect in older children, i.e. bunching of scores at the upper level of an instrument. This might also lead to higher factor loadings in the older age groups.

4.2 Concurrent and predictive validity of the assessment of emotions in moral transgressions

To evaluate the concurrent and predictive validity of the measure, we related anticipated emotions to aggressive and prosocial tendencies and sympathy at two time points. As we found a one-factor structure that held across sex and age, we added the scores of the anticipated emotions across the six scenarios and took the mean as an indicator of children's anticipated emotions following moral transgressions. In line with the results obtained by Malti et al. (2009), we found positive relations between sympathy and children's negative emotions following hypothetical transgressions. Furthermore, prosocial tendencies were weakly related to anticipated negative emotions at the first time point. Surprisingly, no relation was found between self-evaluated emotions and antisocial tendencies. This in contrast with the findings in the meta-analysis of Malti and Krettenauer (2013) where anticipated negative emotions were more related to antisocial than to prosocial tendencies across different age groups. The absent relation between emotions and antisocial tendencies might be due to the way in which the antisocial tendencies of children were assessed. Malti et al. (2009) suggest that observed behaviour is more closely related to anticipated emotions than other-reported

behaviour. Since observed behaviour reflects children's conceptions of sociomoral events it might relate more closely to what children say they would feel in hypothetical situations than would other reported behaviour. Also, a more detailed coding scheme, as well as a measure of intensity of emotions, may provide a stronger relation between emotions following moral transgressions and behaviour (Malti & Krettenauer, 2013). Moreover, data were collected using a one-on-one interview, and childrens' responses may have been influenced by the presence of an adult. Children might have responded more positively than they actually expected their emotions to be after transgressing a moral rule, even though the open answer format reduces the number of social desirable responses (Nunner-Winkler, 2013). Alternative approaches for measuring emotion, such as physiological measures and neuroimaging approaches are less biased. However, self-reports render a more differentiated assessment of emotion than any other method (Pekrun, 2016). We therefore recommend combining multiple types of assessment of anticipated emotions.

Domain variability within the concurrent and predictive validity of the assessment of anticipated emotions in moral transgressions was not thoroughly assessed because of the apparent one factor structure. Further studies could explore whether the relations with sympathy, and pro- and antisocial tendencies hold for specific scenarios and not in others. This would indicate the importance of relating the situations of the moral scenarios to the moral behaviour you want to measure, as Malti and Krettenauer (2013) suggest.

With regard to developmental variability, one interesting result was found when looking into the concurrent and predictive validity of the instrument. In comparison to the anticipated emotions children age 6-8 report, the anticipated emotions of children age 9-10 show a stronger and positive relation with antisocial tendencies. This might reflect social desirable responses of the 9-10 year olds specifically. The younger children might be less inclined to answer in a social desirable way and the older children might be more confident in giving the answer they are thinking of. Another explanation of this unexpected result might again lie in the way in which the antisocial tendencies of children were assessed. Further research could look into the different ways antisocial tendencies can be assessed and how this reflects differences in relations with other constructs, such as anticipated emotions.

4.3 Concluding remarks

Although research on anticipated emotions in moral transgressions has been published for decades, little work used statistical techniques such as factor analysis. This first attempt to look at the reliability and (aspects of) the validity of this type of instrumentation, helped to evaluate the custom for assessment of anticipated emotions in moral transgressions. The instrument appeared to be a reliable one-factor measure of anticipated emotions following hypothetical moral scenarios. The results further reveal some indications for the concurrent and predictive validity of the assessment of emotions in moral transgressions. However, several considerations could optimize the validity of the assessment of anticipated emotions, such as assessing the intensity of the emotions and magnifying the conflicting issues in the scenarios. The potential value of a measure such as this one seems particularly salient in the current educational environment in which there is an increasing awareness to encourage the acquisition of prosocial values and behaviour. More knowledge in the areas of moral developmental research has significant implications for the design of preventive interventions aimed at increasing care, and social justice in children (Malti & Ongley, 2014). It is our hope that the proposed instrument for the assessment of anticipated emotions in moral transgressions can be used to further the valuable applied work and research on this topic.

Chapter 3 Marginal Deviations in Prosocial Behaviour Development: Do Moral Processes Matter?

Abstract

The theory of marginal deviations concerns the processes through which aggressive behaviour that is initially only marginally deviant from the norm gets transformed into higher levels of aggressive behaviour over time (Caprara & Zimbardo, 1996). The current research applies this theoretical framework to prosocial behaviour. We postulated that central processes of moral functioning might help reinforce marginally prosocial children's and discourage marginally nonprosocial children's prosocial behaviour development over time. Specifically, we tested whether sympathy, moral reasoning, negatively valenced moral emotions and morally relevant personality characteristics influence whether marginal deviations in prosocial behaviour develop into higher or lower levels of prosocial behaviour. The results show that marginal deviations in prosocial behaviour do not relate to prosocial behaviour development. However, the results do point to the importance of the accumulation of moral processes for stimulating prosocial behaviour in children. We discuss our findings in relation to the theory of marginal deviations and the role of central moral processes for stimulating the development of prosocial behaviour.

Note. This chapter is based on Jansma, D.J., Malti, T., Opdenakker, M.C.J.L. & Van der Werf, M.P.C. (2017). Marginal Deviations in Prosocial Behaviour Development: Do Moral Processes Matter? *Manuscript submitted for publication.*

1. Introduction

Prosocial behaviour has been defined as voluntary behaviour that benefits others (Eisenberg, Spinrad & Knafo, 2015). Considerable research has tried to explain changes in prosocial behaviour given this behaviour is only moderately stable during the life span (Eisenberg et al., 2015; Koenig, Cicchetti & Rogosch, 2004). Moreover, only by understanding the processes behind changes in prosocial behaviour it is possible to successfully endorse its antecedents (Berkowitz, Sherblom, Bier & Battistich, 2006; Narvaez & Nucci, 2008). Because the seeds of prosocial behaviour emerge in early childhood (Hepach, Vaish, Grossmann & Tomasello, 2016; Malti & Dys, in press), knowledge of the development of prosocial behaviour across childhood is essential to understand and promote prosocial behaviour. This appeals to the increasing awareness of the need for educational systems to encourage the acquisition of prosocial values and behaviour (Brown, Corrigan & Higgins-D'Alessandro, 2012; Fink & Slade, 2016; Reiman & Dotger, 2008; Rupp & Veugelers, 2003).

The theory of marginal deviations aims at explaining the (dis)continuation in children's social behaviour. According to Caprara, Dodge, Pastorelli and Zelli, (2007) marginal deviations in social behaviour have been neglected in past research, which focused mainly on extreme groups or continuous dimensions of behaviour. The theory of marginal deviations argues that marginal deviations in behaviour have the potential to develop into higher levels of the same behaviour. Specifically, the theory concerns the processes through which this development is promoted (Caprara, 1992; Caprara & Pastorelli, 1993; Caprara & Zimbardo, 1996). Even though the theory is assumed to apply to the development of both antisocial and prosocial behaviour, empirical evidence for this proposition is lacking. Until now, empirical studies on the theory of marginal deviations solely focused on aggressive behaviour in children and adolescents (e.g. Caprara, Dodge, Pastorelli & Zelli, 2006; 2007). However, it has been speculated that similar processes are at play in the development of prosocial behaviour (Caprara et al., 2007). This motivated us to study marginal deviations in prosocial behaviour, providing new insights in research and theorizing about prosocial behaviour development.

1.1 The theory of marginal deviations

An important principle of the theory of marginal deviations is that the

processes through which behaviour recurs are particularly present among children who initially display behaviour that is marginally deviant from the norm (Caprara, 1992; Caprara & Pastorelli, 1993; Caprara & Zimbardo, 1996). Marginally deviant behaviour is defined as a small degree of reliable variability in behaviour between a particular child and a normative baseline and has been operationalized as a standard score that is greater than 0 but less than 1 *SD* above the norm, e.g. the mean of initial behaviour (Caprara et al., 2007). In the study of Caprara et al. (2006) marginally aggressive children were classified as those children that initially had a standard score greater than 0 but less than 1 *SD* above the norm of aggressive behaviour. These children were compared to marginally nonaggressive children, initially having a standard score smaller than 0 but more than 1 *SD* below the norm of aggressive behaviour, with regard to their development of aggressive behaviour. Besides that most children's marginally deviant behaviour returned back to the behavioural norm, Caprara et al. (2006) found that in subsequent years marginally aggressive children were more aggressive than marginally nonaggressive children.

In our study, we are interested in prosocial behaviour development in marginally prosocial children as well as marginally nonprosocial children, respectively initially having a standard score greater than 0 and less than 1 *SD* above the norm and smaller than 0 and more than 1 *SD* below the norm of prosocial behaviour. Children that are marginally nonprosocial might show similarities with children that are marginally aggressive and might therefore be more prone to develop lower levels of prosocial behaviour than marginally prosocial children who are more prone to develop higher levels of prosocial behaviour. Given our interest in both marginally prosocial and marginally nonprosocial children, marginally deviant behaviour from now on refers to both positive and negative deviations from the behavioural norm, e.g. the mean of initial behaviour. Marginal deviations in prosocial behaviour will thus be an overarching term referring to both being marginally prosocial and being marginally nonprosocial.

According to the theory of marginal deviations, marginal deviations in behaviour will only lead to higher levels of this behaviour in combination with other factors. Most of the time, children's marginally deviant behaviour returns back to the behavioural norm. However, the factors that influence the transformation of marginal deviations into higher levels of the same behaviour lie in the marginally deviant child, his/her environment and in the interaction between the marginally deviant child and his/her environment (Caprara et al., 2007; Caprara

& Zimbardo, 1996). Let us first discuss the influence of the environment. An environment in which tolerance and cooperation characterize interpersonal exchanges and in which there exists a relatively high consensus on what is (in)appropriate behaviour, a marginally prosocial child might not be noticed and/or perceived as extremely prosocial. In an adverse environment, a marginally prosocial child may, when repeated and aggregated over time, arouse feelings and reactions in others. In turn, these feelings and reactions of others may, when repeated over time, lead to reduction of prosocial behaviour. In a supportive environment, however, the feelings and reactions of others might lead to the enhancement of prosocial behaviour (Stormshak & Webster-Stratton, 1999). For example, peers sometimes respond in a reinforcing manner to prosocial actions and such reinforcement may stimulate children's prosocial behaviour over time (Grusec & Redler, 1980). In this way, a cyclical process may occur in which marginally prosocial children elicit positive (peer) reactions for their behaviour, which in turn encourages prosocial behaviour (Eisenberg et al., 2015). A reverse process of rejection could occur for marginally nonprosocial children. In an supportive environment, marginally nonprosocial children may arouse feeling and reactions in other that, in turn, lead to nonprosocial trait judgments in others. Peers might then respond in a rejecting manner to nonprosocial actions leading to a negative spiral and a discouragement of prosocial behaviour (Caprara & Zimbardo, 1996; Eisenberg et al., 2015).

Psychological processes within the marginally deviant child also influence the development of initial marginal deviations into habitual behaviour. Marginal deviations in behaviour namely not always elicit strong feelings and reactions in others. In most cases, feelings quickly dissipate and the marginally deviant stimulus may be dismissed, reinterpreted, or ignored by observers (Dodge, 2006). In this case the marginal deviant behaviour will be bought back to the norm. Though, when a child's marginal deviations in one behavioural domain are coupled with marginal deviations in other significant domains, the chance increases that an observer notices this deviation. Caprara et al. (2006) demonstrated that the accumulation of risk factors like hyperactivity, cognitive difficulties, and low social preference predicts changes in aggressive behaviour depending on initial marginal aggressive behaviour. The accumulation of risk factors exerted more impact on the development of aggression for marginally aggressive than for marginally nonaggressive children. In a similar fashion, the accumulation of protective factors

might stronger (positively) influence the development of prosocial behaviour among initially marginally prosocial children than among initially marginally nonprosocial children. On the other hand, the absence of the accumulation of protective factors might have a stronger (negative) impact on prosocial behaviour development among initially marginally nonprosocial children than initially marginally prosocial children. Thus, when multiple (marginal) deviations correspond, the environment of the marginally deviant child more easily notices and reacts to the behavioural deviance and hence reinforces the particular deviant behaviour. In this way, personal psychological processes moderate the effect of marginal deviations on the feelings and reactions of others and thereby the potential reinforcement or discouragement of the marginal deviations in behaviour (Caprara et al., 2007).

1.2 Moral processes that can nurture prosocial deviance

Thus far, it remains unexplored whether and which psychological risk and protective factors moderate the effect of marginal deviations in prosocial behaviour on the development of prosocial behaviour. Also, it is unclear whether (the accumulation of) psychological factors plays a role in predicting whether marginal deviations in prosocial behaviour develops into higher or lower levels of prosocial behaviour. Here, we focus on central processes of moral functioning that may predict whether marginal deviations in prosocial behaviour lead to the reinforcement or discouragement of initial prosocial behaviour. Compelling evidence exists showing that prosocial behaviour is positively associated with several aspects of children's moral functioning (Eisenberg et al., 2015). Therefore, central moral processes might also contribute to the development of marginal deviations in prosocial behaviour into higher or lower levels of prosocial behaviour.

Moral functioning refers to the psychological process that a person invokes in order to respond to and resolve a specific problem, conflict or dilemma that requires a moral action (Tappan, 2006). This moral action "bears on the interest or welfare either of society or at least of persons other than the judge or agent" (Gewirth, 1984, p. 978). The whole range of moral functioning can be captured by the Four Component Model of Rest's (1983; 1986), a widely used model postulating the underlying psychological processes preceding a moral action (Myyry, Juujärvi, Pesso, 2010; Vozzola, 2014). These four underlying psychological processes are moral sensitivity, moral reasoning, moral motivation, and moral character. In more

recent terminology these processes more or less translate into sympathy, moral reasoning, negatively valenced moral emotions, and morally relevant personality characteristics (e.g. Ahadi & Rothbart, 1994; Bebeau, Rest & Narvaez, 1999; Cumberland-Li, Eisenberg & Reiser, 2004; Malti & Ongley, 2014; Malti, Gasser & Buchmann, 2009; Mower, Robinson & Vandenberg, 2015; Nunner-Winkler, 2013; Weisberg, DeYoung & Hirsh, 2011). Sympathy, relating to the first component, concerns feelings of concern or sorrow for the other person based on an understanding of that person's circumstances (Eisenberg, 2000). Moral reasoning describes the process in which individuals, using logic and self-reflection, determine why a specific act is right or wrong from a moral perspective (Malti & Ongley, 2014). By attributing negatively valenced (i.e. sadness, guilt) emotions to a wrongdoer, children may indicate the relative importance they attach to moral conformity versus need satisfaction when needs conflict with norms (Nunner-Winkler, 1999; 2007). Lastly, inhibitory control and the personality dimensions agreeableness and conscientiousness can be considered morally relevant personality characteristics (Ahadi & Rothbart, 1994; Graziano & Eisenberg, 1997). Inhibitory control refers to behavioural and cognitive suppression of interferences from the environment (Ahadi & Rothbart, 1994). Conscientiousness traits are related to self-discipline, orderliness and goal pursuit whereas agreeableness has a strong link with the regulation of emotions and traits related to a desire to maintain social harmony (Cumberland-Li, et al., 2004; Weisberg et al., 2011).

Surprisingly, the relations between these four central moral processes and prosocial behaviour have remained relatively unexplored in a simultaneous way (Hardy, 2006; Morton, Worthley, Testerman & Mahoney, 2006). By themselves, however, sympathy, moral reasoning, negatively valenced moral emotions, and morally relevant personality characteristics have repeatedly been found to positively relate to prosocial behaviour in children. For example, positive links between sympathy and prosocial behaviour have been found to exist both within specific contexts and at the dispositional level (Eisenberg et al., 2015; Malti et al, 2016). Also, moral reasoning, the second component, tends to be positively related to prosocial behaviour (e.g. Carlo, Hausmann, Christiansen & Randall., 2003; Carlo, Mestre, Samper, Tur & Armenta, 2011; Eisenberg, Carlo Murphy & Court, 1995). In a meta-analysis of Malti and Krettenauer (2013) small-size relations were found between negatively valenced emotions, the third component, and prosocial behaviour across childhood and adolescence. Lastly, prosocial children tend to have

high inhibitory control (e.g. Beauchaine et al., 2013; Carlo, Crockett, Wolff & Beal, 2012; Laible, Carlo, Panfile, Eye & Parker, 2010; Padilla-Walker & Christensen; 2011). Moreover, the personality of moral exemplars has repeatedly been found to orient toward conscientiousness and agreeableness (Walker, 1999; Walker & Hennig, 2004). Hence, there are indications that sympathy, moral reasoning, negatively valenced moral emotions, and morally relevant personality characteristics could be nurturing factors moderating the effects of marginal deviations in prosocial behaviour on the development of prosocial behaviour.

Additionally, the accumulation of central moral processes could also play a role in the development of prosocial behaviour. According to Rest (1986) the order of the four moral processes is not chronological; the moral components do not precede one another, but are distinct processes that might interact and influence one another. Moreover, all four moral processes are central in order to act in a morally relevant, prosocial manner. Following this argumentation, the accumulation of moral processes might stimulate the development of prosocial behaviour and the absence of moral processes might discourage the development of prosocial behaviour. Also, the accumulation of moral processes or the complete absence of moral processes might strengthen the effects of marginal deviations in prosocial behaviour on the positive or negative development of prosocial behaviour.

1.3 Present Study

The main aim of this study is to obtain insights into the effect of marginal deviations in prosocial behaviour on prosocial behaviour development in middle childhood and to the contribution of moral processes to this effect. Whereas some initially marginally prosocial children may develop higher levels of prosocial behaviour, the majority of them may not, and it is argued that the reasons for this difference in outcomes can be partly sought in both the unique and accumulated effects of individual differences in moral functioning. Additionally, some initially marginally nonprosocial children may develop lower levels of prosocial behaviour, even though the majority may not, also depending on the unique and accumulated effects of individual differences in moral functioning. In accordance with existing research, prosocial behaviour was operationalized using global indices of dispositional prosocial behaviour (Eisenberg et al., 2015).

Even though the theory of marginal deviations is said to apply to the development of both antisocial and prosocial behaviour, as far as we know the

theory has only been empirically tested for aggressive behaviour (e.g. Caprara et al., 2006). However, it is known that prosocial and antisocial behaviour are not the opposite ends of a single dimension (Hawley, Little & Card, 2007). Also, the processes underlying prosocial behaviour might be quite different from processes underlying antisocial behaviour (Baumeister et al., 2001; Krueger, Hicks & Mcgue, 2001; Malti & Krettenauer, 2013; Rothbart & Park, 1986). Moreover, Caprara et al. (2006) hypothesized that the behavioural consequences of being marginally prosocial or marginally antisocial are unlikely to be symmetrical given the higher salience of being marginally antisocial (Baumeister et al., 2001; Rothbart & Park 1986). Following this line of reasoning, being marginally prosocial will elicit less reinforcement than being marginally antisocial and might not as easily lead to higher levels of prosocial behaviour as being marginally antisocial leads to higher levels of antisocial behaviour. Moreover, being marginally nonprosocial might relate to being marginally antisocial and therefore elicit more reinforcement than being marginally prosocial. In this way, being marginally nonprosocial might easier lead to lower levels of prosocial behaviour than being marginally prosocial leads to higher levels of prosocial behaviour. At this moment, however, empirical evidence for this proposition is lacking and therefore our undertaking is rather explorative.

Following the study of Caprara et al. (2007) on aggressive behaviour, we first compared initially marginally prosocial children, having a standard score greater than 0 and less than 1 SD above the norm of prosocial behaviour, with initially marginally nonprosocial children, having a standard score smaller than 0 and more than 1 *SD* below the norm of prosocial behaviour, with regard to the development of prosocial behaviour. Generally, we expected both groups to develop towards the norm, e.g. the mean of initial prosocial behaviour. This means that marginally prosocial children tend to express less prosocial behaviour over time and marginally nonprosocial children tend to express more prosocial behaviour over time.

Second, (the accumulation of) a child's moral functioning is thought to positively develop prosocial behaviour over time for both marginally prosocial children and marginally nonprosocial children. We therefore examined whether sympathy, moral reasoning, negatively valenced moral emotions and morally relevant personality characteristics positively predicted the development of prosocial behaviour for both marginally prosocial and marginally nonprosocial children, on top of the effect of marginal deviations. Moreover, we examined

whether the effect of the aforementioned central moral processes on the development of prosocial behaviour is stronger when moral processes accumulate.

After examining the main effects of marginal deviations in prosocial behaviour and (the accumulation of) moral processes on prosocial behaviour development, we were interested in the moderating effect of central moral processes and their accumulation, on the link between marginal deviations in prosocial behaviour and the development of prosocial behaviour. The theory of marginal deviations assumes that when a child's marginal deviations in prosocial behaviour are coupled with marginal deviations in central moral processes, the chance increases that observers notice the marginal deviations in prosocial behaviour and reinforce it. The occurrence of a moderating effect thus strongly depends on the observers and their reactions to marginal deviations in prosocial behaviour. First of all, when observers do not notice nor reinforce marginal deviations in prosocial behaviour, because of the low salience of marginal deviations in prosocial behaviour, there will be no moderating effect depending on central moral processes. Both the prosocial behaviour of marginally prosocial children and marginally nonprosocial children will then be equally influenced by central moral processes. Second, the reactions of observers could be equally reinforcing for both marginally prosocial and marginally nonprosocial children. In the case that observers equally notice and reinforce being marginally prosocial and being marginally nonprosocial, high scores on (the accumulation of) moral processes are thought to positively influence prosocial behaviour development for marginally prosocial children more than for marginally nonprosocial children. Additionally, low scores on (the accumulation of) moral processes are thought to negatively influence prosocial behaviour development for marginally nonprosocial children more than for marginally prosocial children. Third, given the low salience of prosocial behaviour, being marginally prosocial might stay unnoticed and develop towards the norm. Following this argumentation, the salience of being marginally nonprosocial might be higher than being marginally prosocial. Therefore, being marginally nonprosocial might easier decrease prosocial behaviour than being marginally prosocial increases prosocial behaviour when coupled with (the accumulation of) moral processes. Being marginally nonprosocial combined with low scores on the (accumulation of) moral processes will then discourage the development of prosocial behaviour more than the combination of being marginally

prosocial and high scores on the (accumulation of) moral processes encourages the development of prosocial behaviour.

Lastly, the above processes could operate more readily for certain groups of children than others. In the study of Caprara et al. (2007) marginally aggressive boys were more easily prompted in aggressive growth in combination with risk factors than were marginally aggressive girls. With regard to prosocial behaviour, however, females are expected and believed to be more prosocial than males (Carlo, Mestre, McGinley, Samper, Tur & Sandman, 2012; Malti, Gummerum, Keller, Buchmann, 2009; Nantel-Vivier et al., 2009; Warden, Cheyne, Christie, Fitzpatrick & Reid, 2003). Therefore, since they fit the stereotype, girls could be more susceptible to the influence of marginal deviations in prosocial behaviour on prosocial behaviour development than boys when coupled with (the accumulation of) moral processes. Another relevant question is whether children are especially vulnerable to processes affecting marginal deviations at a particular age. Generally, youth are most susceptible to deviant peer influences during the early adolescent years, when social comparison processes are particularly salient (Dodge & Sherrill, 2006). Since our sample concerns children age 6 to 12, children could more readily develop (or cease) prosocial behaviour based on marginal deviations in prosocial behaviour and (the accumulation of) moral processes when getting older.

2. Method

2.1 Participants

The present study was part of a longitudinal research project examining moral development and prosocial and antisocial behaviour in middle childhood. The study makes use of the data collected at eleven schools located in the Northern part of the Netherlands recruited via the personal network of the researcher. Ethical consent for this study was obtained from the Ethical Committee Pedagogical and Educational Sciences from the University of Groningen. First, school principals and teachers were asked for consent. Parental consent letters were then distributed to obtain permission for their children's participation (acceptance rate: 99%). The data collection took place at the beginning and at the end of one school year: in September/October 2014 and May 2015. During the first measurement occasion the participants of this study were 1080 children aged 6 to 13 years (*M* age = 9.11, *SD* = 1.86, 48% females) divided over 54 classrooms**.** The average number of children per

classroom was 24.0 (*SD*=5.2; range= 14 to 32). The children attended regular education and were predominantly white and of Dutch descent (91.6%), not being representative for the population in the Netherlands (77.3%, CBS, 2017). The second measurement occasion 6 children entered the sample and 7 children left, resulting in a sample of 1079 children with the same aforementioned characteristics.

2.2 Procedure

All children participated in an one-on-one interview assessing moral reasoning and negatively valenced moral emotions and filled in an online questionnaire assessing sympathy, morally relevant personality characteristics and prosocial behaviour. Both were administered by undergraduate students and graduates. The students and graduates administering the interview and questionnaire all received a training of two days and elaborate feedback on their interviews. Children in grade 1 and those with difficulty concentrating and/or reading were assessed one-on-one, where the researcher read out the online questionnaire. Otherwise, the children were seated in groups of four to ten at computers spaced sufficiently to ensure privacy. The children were instructed to provide their own responses to the questions in the questionnaire and interview and were informed that there were no right or wrong answers. Great care was taken to assure that their answers would remain strictly confidential and would not be revealed to anyone. The online questionnaire and the interview took 15-25 minutes each. The interviews were recorded and transcribed afterwards.

2.3 Measures

Prosocial behaviour.

Peer ratings of prosocial behaviour ('Who helps other children?' and 'Who says and does nice things?') were obtained during the first and second measurement occasion. The children were given a roster with the names of their classmates and they could select as many, or as few, classmates as they wanted. The frequency with which each participant was nominated was divided by the number of classmates who were nominating, in order to adjust for class size, and multiplied by a hundred to obtain a percentage (see Coie and Dodge's procedure, 1983). This sociometric method of assessing prosocial behaviour has been used in other studies that established its validity (Carlo, Koller, Eisenberg, Da Silva & Frohlich, 1996).

Cronbach's alpha for prosocial behaviour was .80 for the first measurement occasion and .87 for the second measurement occasion.

Marginal Deviations in Prosocial behaviour. Standardized prosocial nomination scores were then used to classify children as marginally prosocial or marginally nonprosocial. Instead of comparing a child's score on prosocial behaviour to an overall norm, e.g. the mean prosocial behaviour across participants (Caprara et al., 2007), we compared a child's score on prosocial behaviour to the class norm, e.g. the mean prosocial behaviour across classmates. For Dutch children their classmates are the primary venue for experiences with peers and taking part in groups, considering that in the current education system children are normally together with the same classmates for the first eight years of their education. Therefore, we argue that the norm across classmates is more relevant for a child's prosocial actions than the overall norm in the population. In support of this, research shows that processes of social influence among classmates can be rather persuasive for both prosocial and antisocial behaviour of children (Dodge & Sherrill, 2006; Espelage, Holt & Henkel, 2003; Juvonen & Galvan, 2008). Through emergent properties, such as norms and processes, peers help define the type and range of relationships and interactions that are likely or permissible (Rubin, Bukowski & Laursen, 2011). In line with this, marginally prosocial children were those children that obtained a z-score greater than 0 but less than 1 *SD* above the classroom mean of initial prosocial behaviour and marginally nonprosocial children were those children who received a z-score smaller than 0 but more than 1 *SD* below the classroom mean. Because girls would be overrepresented by this criterion, this classification procedure was employed only after standardizing sociometric scores within gender as well (i.e., to minimize the chances that the group of marginally prosocial children included a disproportionate number of girls) (e.g. following Caprara, et al., 2007). This means that children's scores on prosocial behaviour were first standardized relative to the scores of children with the same gender. Then, these scores were standardized relative to the scores of their classmates.

Development of Prosocial behaviour. The same percentage of peer ratings for prosocial behaviour obtained during the first and second measurement occasion were subtracted, indicating the development of prosocial behaviour over the course of one year. Cronbach's alpha for the development of prosocial behaviour was .75.

Moral processes.

The moral processes and covariates were assessed at the first measurement occasion in order to predict prosocial behaviour development.

Sympathy. Sympathy was assessed with eight statements based on the empathic concern measure of Zhou, Valiente, and Eisenberg (2003). After each statement (e.g. "When I see someone being picked on, I feel sorry for them") children were asked whether the sentence described him/her or not, and if so, how strongly on a scale from 0 to 2: "No, this does not sound like me", "This is sort of like me", and "This is really like me" (α=.80).

Moral reasoning. Moral reasoning was assessed by means of an interview using a series of six depicted hypothetical transgressions covering three moral domains: fairness (not winning fairly, not keeping word), victimization (verbal bullying) and omission of prosocial duties (refusing to share pencils, refusing to help someone in pain, refusing to stand up for someone) (Jansma, Malti, Opdenakker & Van der Werf, 2017). In the beginning of each scenario, the children were asked why it was right or wrong to transgress the moral rule. Using a validated coding system (Malti, Gasser & Buchmann, 2009) children's' reasons for their choices were coded as either moral (i.e., those which refer to moral norms and empathic concern for the victim), non-moral, or other/ unclassifiable. Interrater agreement on the coding was K=.87 for winning fairly, K=.91 for keeping word, K=.88 for sharing pencils, K=.94 for helping someone in pain, K=.88 for verbal bullying and K=.83 for not standing up for someone. The reliability of the scale scores as measured with Cronbach's Alfa was .33 (k=6).

Negatively valenced moral emotions (NVME). NVME were assessed using the same interview using six validated moral transgressions. Children were asked how they would feel if they transgressed the moral rule (emotion attribution). By attributing an emotion to a hypothetical wrongdoer, children may indicate the relative importance they attach to moral conformity versus need satisfaction when needs conflict with norms. Following Jansma, Malti, Opdenakker and Van der Werf (2017) anticipated emotions were coded as negative (e.g., bad or half well and half bad) or positive (e.g., happy) emotions. Intercoder reliability of the binary coding of emotions was K=.97 for winning fairly, K= .94 for keeping word, K= .97 for sharing pencils, K= .99 for helping someone in pain, K= .94 for verbal bullying and K= 1.00 for not standing up for someone. The reliability of the scale scores as measured with Cronbach's Alfa was .76 (k=6).

Morally relevant personality characteristics. As described in the introduction, inhibitory control and the personality traits agreeableness and conscientiousness are considered morally relevant personality characteristics. Inhibitory control was assessed with an adjusted and translated version of the subscale "Inhibitory Control" of the Early Adolescent Temperament Questionnaire-Revised (Ellis & Rothbart, 1999). Children reported online (e.g. "I am good at self-discipline.") on a scale ranging from 1 (never) to 5 (almost always) (k=11, α =.62). Teachers ratings of the personality dimensions agreeableness and conscientiousness concerned one item for each personality characteristic ranging from 1 (not agreeable, unconscientious) to 5 (agreeable, conscientious) with more elaborate descriptions of these characteristics. These teacher ratings were derived from the Five-Factor Personality Inventory (FFPI) assessing the Big Five factors of personality (Hendriks, 1997). The five factor scores of the FFPI appeared to be stable and valid in the normal population in the Dutch cohort study COOL 5-18 (Driessen, Mulder, Ledoux, Roeleveld & Van der Veen, 2009). Additionally, the pupil items of the FFPI were found to be highly related to the teacher ratings on the five personality factors (GION & CITO, 2008).

Accumulation of central moral processes. The accumulation of central moral processes is operationalized following Caprara et al. (2006). First, the presence of a moral process was determined if a child's score was above the median of the entire sample (i.e. 1 = presence of moral process, 0 = absence of moral process). Morally relevant personality characteristics were present when a child's score was above the median for agreeableness, conscientiousness, and inhibitory control. Then, each child was assigned an accumulation score indicating whether how many moral processes were present in the child (i.e. the accumulation score ranged from 0 to 4).

Covariates.

Age, gender, socioeconomic status, and scholastic ability. Age, gender, socioeconomic status, and scholastic ability were included as covariates. All have repeatedly been found to be interwoven with children's prosocial behaviour (development) (for gender see Carlo et al., 2012; Malti et al., 2009; Nantel-Vivier et al., 2009;Warden et al., 2003; for age see Eisenberg & Fabes, 1998; for scholastic ability see Eisenberg, Fabes & Spinrad, 2006; and for socioeconomic status see Eisenberg et al., 2006; Rueden, Gosch, Rajmil, Bisegger & Ravens-Sieberer, 2006).

Socioeconomic status of the children was measured by averaging the level of highest completed education of their father and mother on a scale from 1 to 7 (1=kindergarten, 2 primary education, 3= lower levels secondary education 4= higher levels secondary education, 5 = Bachelor, 6 = Master, 7=Master+ in accordance with the standard education grouping of UNESCO, 2006). Scholastic ability was measured by taking the average of nationally normed achievement tests on math and reading developed by the Dutch Central Institute for Test Development (2017). Scores range from 1 (lowest 20%) to 5 (highest 20%).

2.4 Statistical Analysis

The relations between the (accumulation of) moral processes, marginal deviations in prosocial behaviour and the development of prosocial behaviour were examined with a multilevel regression analysis using MLwiN (Rasbash, Charlton, Browne, Healy & Cameron, 2005), adjusting for the covariates socioeconomic status, scholastic ability, gender and age and for initial prosocial behaviour. Since the development of prosocial behaviour might be related to the initial level of prosocial behaviour, we also controlled for initial prosocial behaviour in order to solely predict the development of prosocial behaviour. We carried out two three-level multilevel analyses consisting of lower-level observations (i.e. level 1 units of analysis) nested within higher-level observations (i.e. level 2 units of analysis) nested within even higher-level observations (i.e. level 3 units of analysis) predicting the development of prosocial behaviour. In the context of the present study, children were nested within classrooms within schools. Multilevel analysis takes into account this data structure by using a hierarchical linear model that allows for within group variability as well as between group variability (Snijders & Bosker, 2011). The first multilevel analysis concerns the predictive value of moral processes and marginal deviations in prosocial behaviour for the development of prosocial behaviour. The second multilevel analysis focusses on the predictive value of the accumulation of moral processes and marginal deviations in prosocial behaviour for the development of prosocial behaviour. In both models the moderating effect of the (accumulation of) central moral processes on the effect of marginal deviations in prosocial behaviour on the development of prosocial behaviour was examined.

The predictors of interest were individual-specific and were included in the model as level 1 predictors. There were no predictors at level 2 and 3; the class and

school level. The starting point of the multilevel model was the so-called empty model without any predictors. The empty model provided preliminary information about the variance of the dependent variable, i.e. the development of prosocial behaviour between the first and second measurement occasion, between individuals within classes and schools (i.e. level 1), between classrooms within schools (i.e. level 2) and between schools (i.e. level 3). . Each subsequent model was compared to the preceding one to evaluate whether the inclusion of additional predictors provided a better fit of the data. In order to take a more or less straight way through the jungle of possible multilevel models, three forward steps after the empty model were distinguished: (1) adding fixed predictors at level 1 (i.e. the covariates and initial prosocial behaviour); fixed effects do not vary across classrooms and can be regarded as the average effect over the whole population of children, (2) adding the fixed explanatory variables of main interest at level 1 (i.e. marginal deviations in prosocial behaviour and the moral processes or the accumulation of moral processes) to evaluate the unique role in the prediction of the dependent variable while controlling for the variables entered in the previous model step, (3) adding the interaction effects between (accumulation of) moral processes and marginal deviations in prosocial behaviour at level 1 (i.e. moral process*marginal deviations or accumulation*marginal deviations), and (4) adding the interaction effects between age, sex and the interaction effect between (accumulation of) moral processes and marginal deviations in prosocial behaviour at level 1 (i.e. moral processes*marginal deviations*age and moral processes*marginal deviations*sex, or accumulation*marginal deviations*age and accumulation*marginal deviations*sex) and their underlying interaction effects (i.e. moral processes*age and moral processes*sex and marginal deviations*age and marginal deviations*sex). In addition to the linear effect of the accumulation of moral processes, a quadratic effect was added to the model to reflect the acceleration of the impact of the moral processes. Having tested for random slopes and included those that were significant ($p<0.05$), the significance of both the fixed and interaction effects was evaluated with the t-test, based on the ratio of parameter estimate to standard error. Insignificant random slopes were removed from the model and this resulted in the final model. Comparisons between the deviance statistic of the final models and the deviance statistic of models with less parameters indicate that the final models we present were significantly the best fitting models

for the given data. For ease of interpretation as well as estimation, the predictors were centred around the grand mean prior to statistical analysis.

3. Results

3.1 Missing data

Unfortunately, not all teachers filled in the questionnaire about the morally relevant personality characteristics agreeableness and conscientiousness of the children in their class leading to a reduction of the number of children that could be included in the present study. Furthermore, the remaining data contained some missing values, mainly due to unit non-response or schools not being able to provide background information on parental education or scholastic ability. The 680 children with complete data did not differ from the other children in terms of age, gender, moral reasoning, moral emotions, conscientiousness and inhibitory control ($t(1084)=.41$; $p=.68$; $\chi^2(1)=.24$; $p=.63$; $t(954)=-1.48$; $p=.14$; $t(954)=-1.03$; $p=.30$; $t(775)=-.69$; $p=.49$; $t(1033)=-.44$; $p=.66$). However, non-participating children had a lower mean score on prosocial behaviour, agreeableness, sympathy, scholastic ability and socioeconomic status than participating children (respectively $t(1078)=-7.69$, $p<.01$, $d=.48$; $t(774)=-2.35$; $p=.02$, $d=.26$; $t(1033)=-5.09$, $p<.01$, $d=.32$; $t(1041)=-3.40$, $p<.01$, $d=.07$; $t(1042)=-3.02$; $p<.01$, $d=.20$) and a higher mean score of the development of prosocial behaviour ($t(1071)=2.31$, $p=.02$, $d=.15$). In order to compute the multilevel regression model in MLwiN, missing data was deleted in a listwise manner. Considering that our data is missing not at random, bias in analyses based on multiple imputation may be as big as or bigger than the bias in analyses of only complete cases (e.g. Sterne et al., 2009).

3.2 Descriptive statistics

In total, 267 children (32,9%) were classified as marginally prosocial (116 boys and 151 girls) and 224 children (39.6%) were classified as marginally nonprosocial (127 boys and 97 girls). The other 187 children (27,5%) scored higher than 1 *SD* or lower than -1 *SD* above the class norm of prosocial behaviour and 2 children scored on the norm. The means and standard deviations for all measures are presented in Table 1 at each time point for the marginally prosocial children, the marginally nonprosocial children, and the entire sample (N=680). Within the entire sample 48.8% of the children were girls, within the marginally nonprosocial

children 43.3% were girls and within the marginally prosocial children 56.9% were girls. The marginally prosocial children had higher scores on all moral processes compared to the marginally nonprosocial children. Also, the mean accumulation of moral processes was higher for marginally prosocial than marginally nonprosocial children. Logically, marginally prosocial children were initially more prosocial than marginally nonprosocial children. Also, marginally prosocial children showed less development of prosocial behaviour than marginally nonprosocial children.

Table 2 presents correlations between the covariates, the central moral processes and the development of prosocial behaviour for marginally prosocial and marginally nonprosocial children. Although the development of prosocial behaviour was significantly negatively related to initial prosocial behaviour, the correlation was modest in magnitude. This indicates that the development of prosocial behaviour is accompanied by slightly lower initial prosocial behaviour, or that children with higher levels of initial prosocial behaviour show slightly less development or a decline in prosocial behaviour. Additionally, none of the central moral processes was significantly or meaningfully correlated with the development

Table 1

Mean and standard deviations of the outcome and predictor variables (N=680)

	Entire sample M (SD)	Marginally nonprosocial M (SD)	Marginally prosocial M (SD)
Age	9.10 (1.85)	8.99 (1.99)	9.07 (1.79)
Scholastic ability	3.47 (1.09)	3.33 (1.05)	3.58 (1.04)
Socioeconomic status	4.28 (.86)	4.36 (.84)	4.23 (.83)
Sympathy	2.15 (.47)	2.07 (.47)	2.20 (.47)
Moral reasoning	1.81 (.18)	1.78 (.20)	1.84 (.15)
Moral emotions	1.80 (.27)	1.76 (.29)	1.82 (.27)
Agreeableness	3.65 (1.00)	3.61 (1.06)	3.75 (.91)
Conscientiousness	3.20 (1.07)	3.04 (1.09)	3.32 (.99)
Inhibitory control	3.94 (.49)	3.88 (.53)	3.97 (.47)
Accumulation score	1.91 (.98)	1.73 (.96)	2.10 (.93)
Initial prosocial behaviour	47.70 (17.98)	38.07 (13.27)	55.12 (13.20)
Prosocial behaviour development	1.83 (15.75)	4.33 (14.56)	-.26 (16.30)

of prosocial behaviour. This was true for both marginally prosocial and marginally nonprosocial children. Moreover, the positive correlation between the accumulation of moral processes and the development of prosocial behaviour was small and not significant for both marginally prosocial and nonprosocial children.

Table 2 further shows that both the central moral processes and their accumulation did not moderate the relations between marginal deviations in prosocial behaviour and the development of prosocial behaviour, i.e. the correlations between (the accumulation of) central moral processes and the development of prosocial behaviour are quite similar for marginally prosocial and marginally nonprosocial children.

Table 2

Correlations between covariates and moral processes and the development of prosocial behaviour for marginally prosocial and marginally nonprosocial children (n=491)

	Marginally nonprosocial	Marginally prosocial
Age	.08	.12
Scholastic ability	.08	.07
Socioeconomic status	-.05	-.01
Sympathy	-.08	.04
Moral reasoning	.08	.07
Moral emotions	.04	.05
Agreeableness	.05	.03
Conscientiousness	.11	-.02
Inhibitory control	-.13	-.05
Accumulation score	.06	.09
Initial prosocial behaviour	-.18**	-.18**

3.3 Multilevel analyses

In the following section the results of two different multilevel analyses will be discussed. The first multilevel analysis concerns the predictive value of moral processes and marginal deviations in prosocial behaviour on the development of prosocial behaviour. The second multilevel analysis focusses on the predictive value of the accumulation of moral processes and marginal deviations in prosocial behaviour for the development of prosocial behaviour. In both models the

moderating effect of the (accumulation of) central moral processes on the effect of marginal deviations in prosocial behaviour on the development of prosocial behaviour was examined. Interaction terms linking age and gender to this interaction between moral processes and marginal deviations in prosocial behaviour were inserted as well. Only the marginally prosocial and marginally nonprosocial children were incorporated in the analysis. For reasons of clarity, random slopes and underlying interaction effects can be found in Appendix D.

3.3.1 The moderating effects of moral processes on the relation between marginal deviations in prosocial behaviour and prosocial behaviour development

Table 3 summarizes the results of the first multilevel analysis testing the predictive value of moral processes and marginal deviations in prosocial behaviour on the development of prosocial behaviour after controlling for the covariates and initial prosocial behaviour. See Appendix D for the complete model. The multilevel analysis shows that scholastic ability and age positively relate to the development of prosocial behaviour; increased ability and increased age correspond to an increased development of prosocial behaviour (respectively $t(490)=3.59$; $p<.01$, $t(490)=2.21$; $p<.05$). Additionally, girls scored higher on the development of prosocial behaviour

Table 3

Multilevel regression analysis predicting the development of prosocial behaviour (n=491)

	Central moral processes model		Accumulation model	
Fixed effects	*Estimate*	*S.E.*	*Estimate*	*S.E.*
Intercept	-.628	2.254	-.899	2.188
Level 1 (individual) variables				
Girl	6.046*	2.344	5.929*	2.188
Age	1.649*	.746	1.933*	.668
Socioeconomic status	.888	.791	.647	.785
Scholastic ability	2.381*	.664	2.180*	.678
Initial prosocial behaviour	-.392*	.084	-.423*	.085
Marginal deviations in prosocial behaviour	.182	2.115	1.158	1.996
Sympathy	2.108	2.461		
Moral reasoning	-2.994	7.820		
Moral emotions	1.959	5.105		

Agreeableness	.046	1.291		
Conscientiousness	-.314	1.210		
Inhibitory control	-3.322	2.262		
Accumulation score			4.244	2.834
Accumulation score^2			-1.063	.768
Sympathy*Marginal deviations	2.779	3.341		
Moral reasoning*Marginal deviations	-6.655	9.615		
Moral emotions *Marginal deviations	4.131	6.796		
Agreeableness*Marginal deviations	.099	1.659		
Conscientiousness*Marginal deviations	-1.064	1.682		
Inhibitory control *Marginal deviations	-2.629	3.079		
Accumulation score* Marginal deviations			1.278	5.741
Accumulation score^2* Marginal deviations			-.754	1.575
Sympathy*Marginal deviations*Girl	.163	4.769		
Moral reasoning*Marginal deviations*Girl	6.611	12.839		
Moral emotions *Marginal deviations*Girl	1.171	8.298		
Agreeableness*Marginal deviations*Girl	1.437	2.333		
Conscientiousness*Marginal deviations*Girl	-.438	2.307		
Inhibitory control *Marginal deviations*Girl	1.018	4.407		
Accumulation score * Marginal deviations*Girl			3.364	7.715
Accumulation score^2* Marginal deviations*Girl			-.039	2.027
Sympathy*Marginal deviations*Age	-.245	1.236		
Moral reasoning*Marginal deviations*Age	2.789	3.472		
Moral emotions *Marginal deviations*Age	-6.752*	2.158		
Agreeableness*Marginal deviations*Age	1.173	.644		
Conscientiousness*Marginal deviations*Age	1.005	.637		
Inhibitory control *Marginal deviations*Age	.438	1.177		
Accumulation score* Marginal deviations*Age			1.353	1.986
Accumulation score^2* Marginal deviations*Age			.126	.531
Random effects	*Var. Comp.*	*S.E.*	*Var. Comp.*	*S.E.*
School level variance	0.000	0.000	0.000	0.000
Class level variance	118.494	33.577	125.526	35.594
Individual level variance	114.665	8.080	119.592	8.461
Deviance	3857.29		3863.94	

* Significant at α= 0.05

than boys (t(490)=2.58; p<.01). As was already apparent from the descriptive results, initial prosocial behaviour was negatively related to the development of prosocial behaviour, meaning that higher initial levels of prosocial behaviours correspond to a decreasing development or even a decline of prosocial behaviour over time, and vice versa (t(490)=-4.67; p<.01).

No significant relation was found between the development of prosocial behaviour and marginal deviations in prosocial behaviour (t(490)=.09; p>.05). Also, no significant relations were found between the development of prosocial behaviour and one of the moral processes. Moreover, the interaction terms between the moral processes and marginal deviations in prosocial behaviour did not reach significance. This indicates that the moral processes do not influence the relations between marginal deviations in prosocial behaviour and the development of prosocial behaviour.

We further examined whether the above moderating processes could operate more readily for certain groups of children. The results show that the links between marginal deviations in prosocial behaviour and the moral processes did not differ according to gender when predicting the development of prosocial behaviour. However, a significant interaction effect was found between age, negatively valenced moral emotions and marginal deviations in prosocial behaviour (t(490)=-3.13; p<.01). The relations are depicted in Figure 1 and were such that moral emotions show a positive effect on the development of prosocial behaviour only for marginally prosocial and old marginally nonprosocial children. For young marginally nonprosocial children moral emotions had a negative effect on the development of in prosocial behaviour. Moreover, the effect of moral emotions on prosocial behaviour development for marginally prosocial children decreases with age. The underlying interaction effect of age*moral emotions was also significant (t(490)=2.70; p<.05) meaning that moral emotions positively relate to the development of prosocial behaviour as children get older.

3.3.2 The moderating effects of the accumulation of moral processes on the relation between marginal deviations in prosocial behaviour and prosocial behaviour development

Table 3 also summarizes the results of the second multilevel analysis testing the predictive value of the accumulation of moral processes and marginal deviations in prosocial behaviour on the development of prosocial behaviour after

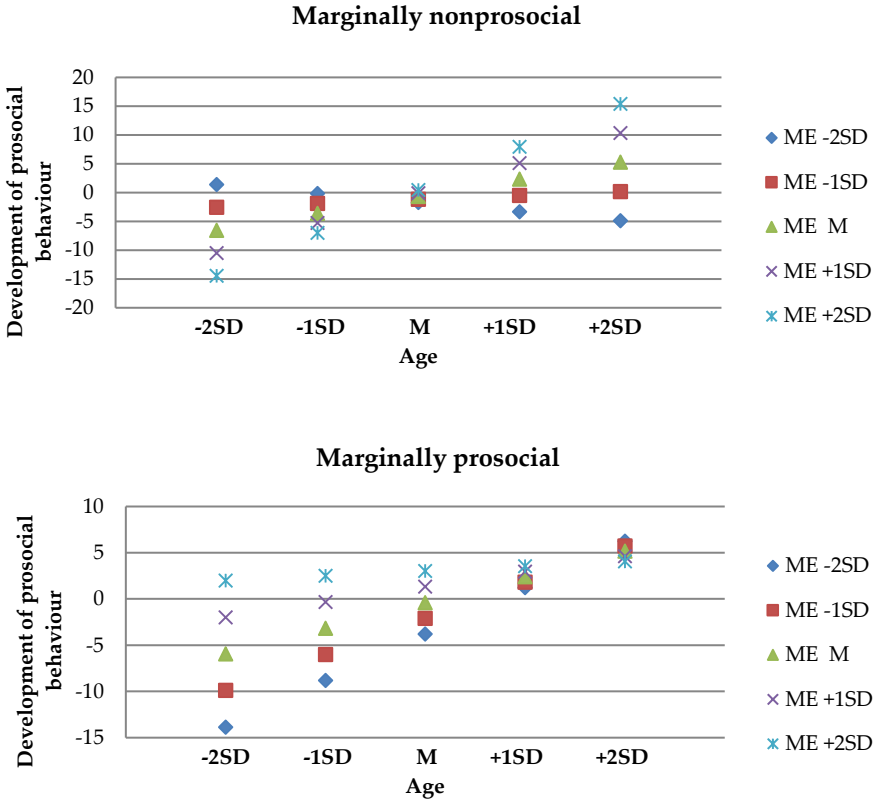


Figure 1. Interaction effect predicting the development of prosocial behaviour between age, moral emotions and marginal deviations in prosocial behaviour.

controlling for the covariates and initial prosocial behaviour. See Appendix D for the complete model. Scholastic ability, age, gender, initial prosocial behaviour and marginal deviations in prosocial behaviour predicted the development of prosocial behaviour similarly to the previous multilevel regression model (respectively $t(490)=3.22$; $p<.01$, $t(490)=2.89$; $p<.01$, $t(490)=2.71$; $p<.05$, $t(490)=-4.79$; $p<.01$, $t(490)=.58$; $p>.05$). Increased scholastic ability and age correspond to an increase in prosocial behaviour and girls showed a greater development of prosocial behaviour than boys. Again, initial prosocial behaviour was negatively related to the development of prosocial behaviour.

No significant relation was found between the development of prosocial behaviour and marginal deviations in prosocial behaviour. The positive linear and

negative quadratic relation between the development of prosocial behaviour and accumulation of moral processes was marginally significant (t(490)=1.50; p=.06, t(490)=-1.38; p=.08). The quadratic relation was added to the model to reflect the acceleration of the impact of the moral processes. The addition of the quadratic accumulation of moral processes did not influence the other effects in the model (except for the linear effect of the accumulation of moral processes). The combination of positive linear and negative quadratic relations indicates that the accumulation of moral processes predicts the development of prosocial behaviour. Moreover, the predictive value is highest when the number of moral components grows from zero onwards and slows down when reaching the presence of all moral processes.

Additionally, the interaction terms between the accumulation of moral processes and marginal deviations in prosocial behaviour were not significant. We already found no relation between marginal deviations in prosocial behaviour and the development of prosocial behaviour. Variations in the accumulation of moral processes do not change this finding.

Last, the links between marginal deviations in prosocial behaviour and the accumulation of moral processes when predicting the development of prosocial behaviour did not differ according to gender or age. This means that the interaction effect between the accumulation score and marginal deviations in prosocial behaviour was insignificant for boys and girls and across age.

4. Discussion

Marginal deviations in prosocial behaviour have been neglected in past research, which focused mainly on extreme groups or continuous dimensions of behaviour (Caprara et al., 2007). The theory of marginal deviations argues that marginal deviations in behaviour also have the potential to develop into higher or lower levels of prosocial behaviour. Our study is among the first to study marginal deviations in prosocial behaviour, providing new insights in research and theorizing about prosocial development. Because the seeds of prosocial behaviour emerge in early childhood (Hepach et al., 2016; Malti & Dys, in press), knowledge of the development of prosocial behaviour in these years is essential to understand and promote prosocial tendencies. We tested whether the development of prosocial behaviour could be predicted by marginal deviations in prosocial behaviour.

Moreover, based on the Four Component Theory (Rest 1983; 1986), we considered the unique and aggregated effects of individual differences in underlying moral processes to be relevant for the development of prosocial behaviour.

Contrary to the theory of marginal deviations, the empirical findings of this study do not show that marginal deviations in prosocial behaviour have the potential to influence the development of prosocial behaviour over time. The positive development of prosocial behaviour for marginally prosocial children only applied to younger children in the sample, and only in combination with higher scores of negatively valenced moral emotions. The negative development of prosocial behaviour for marginally nonprosocial children only applied to younger children in the sample, and only in combination with lower scores of negatively valenced moral emotions. For older marginally nonprosocial children negatively valenced moral emotions had a positive effect on the development of prosocial behaviour. This might be a spurious effect due to multiple comparisons, but it might also indicate that children would more readily develop (or cease) prosocial behaviour based on marginal deviations in prosocial behaviour when getting older. This confirms that children could be susceptible to deviant peer influences during the early adolescent years, when social comparison processes are particularly salient (Dodge & Sherrill, 2006).

Generally, our results suggest that marginal deviations and reinforcing processes leading to aggression do not characterize children who initially show marginal deviations in prosocial behaviour. The relation between marginal deviations in prosocial behaviour and the development of prosocial behaviour was slightly positive, but negligible. Moreover, the interaction effects between the (accumulation of) moral processes and marginal deviations in prosocial behaviour were not apparent from the data. Thus, the relation between marginal deviations in prosocial behaviour and the development of prosocial behaviour did not change due to the (accumulation of) moral processes. A possible explanation for this result is that the behavioural consequences of marginal deviations in prosocial behaviour differ from marginal deviations in antisocial behaviour. We already argued that marginal deviations in prosocial behaviour more often stay unnoticed than marginal deviations in antisocial behaviour. In this way marginal deviations in prosocial behaviour elicit less reinforcement than marginal deviations in antisocial behaviour, or no reaction at all (Baumeister et al., 2001; Rothbart & Park 1986), and may not as easily lead to higher or lower levels of prosocial behaviour. Another

explanation for this result is that the (accumulation of) moral processes do not contribute to the development of marginal deviations in prosocial behaviour into habitual levels of this behaviour. Our study was the first to explore whether and which potential psychological protective factors related to moral processes moderate the effect of marginal deviations in prosocial behaviour on the development of prosocial behaviour. It might be that other psychological risk or protective factors are more important to prosocial behaviour development than the (accumulation of) moral processes.

An interesting addition to the previous result is that initial prosocial behaviour was found to negatively relate to the development of prosocial behaviour for marginally prosocial and marginally nonprosocial children. Further inspection of the data showed that children with high versus low levels of low initial prosocial behaviour tended to develop towards the mean of prosocial behaviour over time. Thus, strong marginal deviations in prosocial behaviour appeared to be compensated by actions in the opposite direction or by not repeating the same action (Dodge, 2006). We would like to speculate that this phenomenon of compensation might be due to processes of social influence among classmates. Since social influence can be rather persuasive for the behaviour of children (Dodge & Sherrill, 2006; Espelage et al., 2003; Juvonen & Galvan, 2008), peer groups help define the type and range of relationships and interactions that are likely or permissible (Rubin, Bukowski & Laursen, 2011). Instead of the hypothesized cyclical process in which prosocial children elicit more positive reactions for prosocial behaviour, which in turn increases their prosocial behaviour, it seems as if marginal deviations from the norm of prosocial behaviour elicits none or negative reactions, which in turn moves children's' marginal deviations in prosocial behaviour towards the norm. It could well be that marginal deviations in prosocial behaviour elicit no reactions in peers. Again, this is in line with the hypothesis that marginal deviations in prosocial behaviour elicit less reinforcement than marginal deviations in antisocial behaviour, because it is socially more expected (Baumeister et al., 2001; Rothbart & Park 1986), and may therefore not as easily lead to higher levels of this behaviour. Another explanation of the negative relation between initial prosocial behaviour and the development of prosocial behaviour is regression to the mean. This is the phenomenon that if a score is extreme on its first measurement, it will tend to be closer to the average on its second measurement. The phenomenon occurs because scores are determined in part by chance. However, due to peer

nominations of prosocial behaviour, the chance of an error occurring due to a single reporter's experience reduces significantly (Rubin, Bukowski & Laursen, 2011).

Contrary to our expectations, sympathy, moral reasoning, and morally relevant personality characteristics did not significantly predict the development of prosocial behaviour for both marginally prosocial and marginally nonprosocial children. This contrasts the theoretical considerations of the current study since moral functioning was supposed to bring advantages in prosocial behaviour. A possible explanation for this result lies in the selective sample of marginally deviant and marginally nondeviant children. In this way the extreme values on prosocial behaviour were left out of the analyses and this might have weakened the relations between moral processes and the development of prosocial behaviour. Another explanation might be that the relations between moral processes and prosocial behaviour vary with the type of prosocial behaviour. Moral functioning more frequently has been associated with prosocial actions that incur a cost than with those low in cost (Eisenberg et al., 2015). Since we used global indices of dispositional prosocial behaviour, the costs of the behaviour were not specified. It might be that children more frequently associated helping and being kind with actions low in costs. Additionally, however, the positive linear and negative quadratic relation between the development of prosocial behaviour and accumulation of moral processes was stronger than the relations between the development of prosocial behaviour and the separate moral components. Thus, there is some indication that the accumulation of moral processes might predict the development of prosocial behaviour of marginally prosocial and marginally nonprosocial children. This preliminary finding suggests that in order to develop prosocial behaviour, it is beneficial to score high on more moral processes and not just one. This is in line with Rest's (1983; 1986) Four Component Model. According to this model, all moral processes must be in place in order to act moral.

An interesting interaction effect was found between age, negatively valenced moral emotions and marginal deviations in prosocial behaviour. Negatively valenced moral emotions positively predicted the development of prosocial behaviour for younger marginally prosocial children and older marginally nonprosocial children. This is in line with the meta-analysis of Malti & Krettenauer (2013). However, for younger marginally nonprosocial children, moral emotions were negatively associated with prosocial behaviour development. This might be a spurious effect due to multiple comparisons, but might also be due to other

mechanisms. For example, it is known that the occurrence of negatively valenced moral emotions show an increase after age 6-7 years due to cognitive development (Arsenio et al., 2006; Krettenauer, Malti & Sokol, 2008). However, it could also be that the young marginally nonprosocial children who score low on moral emotions show prosocial behaviour development because of cognitive development and their growing ability to take the perspective of others. Additionally, the finding that moral emotions matter for prosocial development for younger children that are marginally prosocial (and not for older children) might be explained by the development of moral emotions as well as age related differences in social desirability. Also, it might be that prosocial behaviour is more encouraged and accepted among children age 6-8 than children age 10-12. High moral emotions might therefore more easily tempt the increase of prosocial behaviour when the child is younger and marginally prosocial. In any case, research on moral emotions could be used to a greater degree than in the past to inform our understanding of prosocial behaviour (e.g. Eisenberg et al., 2016).

Several limitations of this study should be mentioned. First, the causal mechanisms underlying the relation between marginal deviations in prosocial behaviour, the (accumulation of) moral processes and the development of prosocial behaviour remain unclear. Even though we made use of longitudinal data, we only made use of two time point that were half a year apart. A lot of change could have happened within or after these six months. Furthermore, we did not look at the specific interaction processes leading marginal deviations in prosocial behaviour to develop into higher or lower levels of prosocial behaviour. We hypothesized an environment in which tolerance and cooperation characterize interpersonal exchanges and in which there exists a relatively high consensus on what is (in)appropriate behaviour, marginally prosocial and marginally nonprosocial children might not be noticed and/or reinforced in their behaviour. However, we did not empirically examine this assumption. Future studies are recommended to examine the reactions of children to prosocial acts and their effect on future prosocial behaviour in order to test the claims underlying the theory of marginal deviations. Second, the data in our study was based on self- and other- ratings, and may be subject to social desirability. Observational measures would be a welcome addition to this type of research. Also, more reliable measures of moral reasoning, moral emotions and inhibitory control might yield additional knowledge on the

relation between marginal deviations in prosocial behaviour, the (accumulation of) moral processes and the development of prosocial behaviour. Third, the definition of marginal deviations remains open for debate and empirical testing. Defining marginal deviations in behaviour as having a standard score greater than 0 but less than 1 *SD* above or smaller than 0 but greater than 1 *SD* under the norm of initial prosocial behaviour is relatively arbitrary. First of all the cut-off at 1 *SD* could be questioned. Second, assigning the label of marginally deviant to a child that scores close to the norm (i.e. the mean) might not reflect true (marginal) deviations with regard to a certain type of behaviour. Another option might be to define marginally deviant as ranging from 0.5 *SD* to 1 *SD*. Future research might try to compare different cut-off scores and establish a good rule for this. Furthermore, we argued that the norm across classmates is more relevant for a child's prosocial actions than the overall norm in the population. We consider this a strength of our study, because the norm across classmates is more relevant for a child's prosocial actions than the overall norm in the population (e.g. Dodge & Sherrill, 2006; Espelage, Holt & Henkel, 2003; Juvonen & Galvan, 2008). However, it might have led to the absence of a clear norm of prosocial behaviour in our sample. There were only two children out of 491 who scored on the norm. Therefore, the assumption of the theory of marginal deviations that there is a clear norm and that marginal deviations from this norm are relevant, is fragile in our study. Lastly, our study is not representative for all children in the Netherlands. A replication would increase insights about the applicability of the theory of marginal deviations to prosocial behaviour and the role of the (accumulation of) moral processes in the development of prosocial behaviour through marginal deviations in this behaviour.

Despite these limitations, the current study provided insights into theorizing about prosocial behaviour and its development. Contrary to the presumptions of the theory of marginal deviations, the empirical findings of this study do not show that marginal deviations in prosocial behaviour have the potential to develop into higher or lower levels of prosocial behaviour over time. In contrast to aggression, prosocial behaviour therefore seems to develop according to compensation rather than the amplification of marginally deviant behaviour. This reminds us that prosocial and antisocial behaviour are conceptually distinct (e.g. Hawley et al., 2007). The findings further suggest that in order for prosocial behaviour to develop, moral processes might play an important role when they accumulate. Also, negatively valenced moral emotions affected prosocial behaviour

development depending on age and marginal deviations in prosocial behaviour. Both these insights are relevant for (educational) interventions aimed at affecting prosocial development. Accordingly, this study contributed towards enhancing our understanding of the role of moral processes in predicting the development of prosocial behaviour. Only by better understanding the underlying processes of prosocial behaviour we might be able to steer them into a prosocial career (Domitrovich, Durlak, Staley & Weissberg, 2017; Malti, Chaparro, Zuffiano, Colasante, 2016).

Chapter 4 Individual and Class Moral Correlates of Children's Behaviour in Bullying Situations

Abstract

The aim of this study was to obtain new insights into the relative contribution of differential moral components to bullying as a group process. The participant roles of bully, assistant, defender, outsider and victim were compared with regard to moral sensitivity, moral reasoning, moral motivation, and moral character at both the individual and class level. Participants were 1,258 children aged 6 to 13 years (49% females) divided over 54 classrooms attending 11 Dutch primary schools. Multinomial multilevel analysis showed that, at the individual level, higher scores on moral reasoning increased the chance being a bully or assistant and decreased the chance of being victimized compared to being an outsider. Higher agreeableness, an aspect of moral character, significantly increased the chance of belonging to the outsider role in comparison to the bully, assistant or defending role. Sympathy increased the chance of being a defender instead of outsider. At the class level, higher class moral emotions and class inhibitory control respectively increased and decreased the chance of being a bully versus being an outsider. These results extend previous research by demonstrating the role of processes at the class level and have potential implications for interventions.

Note. This chapter is based on Jansma, D.J., Opdenakker, M.C.J.L., Malti, T. & Van der Werf, M.P.C. (2017). Individual and Class Moral Correlates of Children's Behaviour in Bullying Situations. *Manuscript submitted for publication.*

1. Introduction

Bullying is a well-known reflection of antisocial behaviour in education prevailing among 15% of primary school children (Veenstra et al., 2005). Bullying occurs in a context of an imbalance of power and is characterized by a repetition of negative actions towards a peer, with the intention to hurt (Olweus, 1993). These actions may include physical or verbal aggression and relational harassment which harms others by means of social manipulation, exclusion or malicious rumours (Camodeca & Goossens, 2005). Bullying in school is not only a great burden for teachers and classroom climate, but also for the perpetrators, victims and society, both financially and socially (Hawker & Boulton, 2000; Nansel et al., 2004; Nishina & Juvonen, 2005; Salmivalli & Isaacs, 2005; Soepboer, Veenstra & Verhulst, 2006). Given the severe consequences of bullying, there is an increasing awareness of the need for educational systems to discourage the acquisition of antisocial values and behaviour (Rupp & Veugelers, 2003). In practice, this awareness is reflected by numerous educational and intervention programs aimed at affecting bullying, prosocial learning and moral functioning (Reiman & Dotger, 2008; Smith, Ananiadou & Cowie, 2003; Smith, Cousins & Stewart, 2005). However, the effectiveness of these programs is generally mixed (Wienke et al., 2014; Willems et al., 2012). Most meta-analyses of bullying programs show small to moderate effect sizes at best. Moreover, positive effects are more likely to be effects on attitudes, knowledge, and perceptions, rather than effects on bullying behaviour (Rivara & Le Menestrel, 2016).

In order to develop effective anti-bullying programs, researchers have attempted to identify specific success-promoting factors of these programs. The current study will focus on moral functioning as a factor being conceptually intertwined with bullying (Cuevas, 2011; Gasser & Keller, 2009) and defending behaviour (Caravita, Gini & Pozzoli, 2012; Menesini & Camodeca, 2008; Pozzoli & Gini, 2010). As such moral functioning might serve as a central process underlying children's bullying. First of all, both bullying and defending are prime examples of morally relevant behaviour in middle childhood, because of its direct effect on the welfare of victims (Turiel, 1983; 1998). Second, different facets of moral functioning have been found to be important to understand individual differences in engagement in bullying situations (Gini, 2006; Hymel, Rocke-Henderson & Bonanno, 2005; Perren & Gutzwiller-Helfenfinger, 2012). The apparent theoretical

and empirical divisions between the study of children's behaviour in bullying situations and their moral correlates seems odd given that both fields share a common focus on the intentional victimization of others (Arsenio & Lemerise, 2001; 2004; Guerra, Nucci & Huesmann, 1994). Therefore, researchers have recently begun to address the role of moral factors in the process of bullying.

Unfortunately, it still remains unclear how different components of children's morality simultaneously affect bullying behaviour (Gasser, Malti & Gutzwiller-Helfenfinger, 2012). Traditionally, research on moral functioning has focused on children's moral reasoning, i.e. their justifications for giving a particular moral judgment. Yet, it has been recognized that a comprehensive account of (im)moral behaviour should not only include cognitive processes but emotional and self-related processes as well (Arsenio & Lemerise, 2004; Olthof, 2010). Rest's (1983; 1986) Four Component Model, grounded in a review of psychological research, postulates that four key psychological components contribute to morally relevant behaviour. These are moral sensitivity, moral reasoning, moral motivation, and moral character. Moral sensitivity concerns interpreting a situation in terms of how people's welfare is affected by possible actions of the subject (Rest, 1983). Sympathy, i.e. feelings of concern or sorrow for the other person based on an understanding of that person's circumstances (Zhou, Valiente & Eisenberg, 2003), is an important element of moral sensitivity (Bebeau, Rest & Narvaez, 1999; Mower, Robinson & Vandenberg, 2015). Moral reasoning regards integrating various considerations to determine what ought to be done. Moral motivation concerns the importance people give to moral values (doing what is right) relative to other values (i.e. self-actualization) (Rest, 1983). Moral character is the ability to persist in a moral task in the face of obstacles (Rest, 1986; 1994). It is expressed by the personality dimensions agreeableness and conscientiousness and by inhibitory control, i.e. behavioural and cognitive suppression of interferences from the environment (Ahadi & Rothbart, 1994; Graziano & Eisenberg, 1997). Conscientiousness concerns traits related to self-discipline, orderliness and goal pursuit whereas agreeableness has a strong link with the regulation of emotions and constitutes of traits related to a desire to maintain social harmony (Ahadi & Rothbart, 1994; Cumberland-Li, Eisenberg & Reiser, 2004; Weisberg, DeYoung & Hirsh, 2011).

A major limitation of research relating moral components to children's behaviour in bullying situations has been the neglect of the complex group nature of bullying in schools (Barchia & Bussey, 2011; Faris & Ennet, 2012; Huitsing &

Veenstra, 2012; Salmivalli, 2010). The result of the neglect of the social nature of bullying in relation to moral components is twofold. First of all, studies have rarely examined the differential associations between moral components and different roles in the bullying process. It is well known that all children in a class are somehow involved in, or aware of, the bullying process, even if they do not actively attack the victim (Atlas & Pepler, 1998; Hawkins, Pepler & Craig, 2001). Specifically, Salmivalli, Lagerpetz, Björkvist, Österman and Kaukiainen (1996) and Salmivalli, Lappalainen and Lagerspetz (1998) identified five participant roles involved in the bullying process, next to the victim: (ringleader) bully, who starts the bullying; assistant, who joins in the bullying, but does not start it; reinforcer, who encourages the bully; defender, who supports the victim; and outsider, who keeps out of the bullying situation. Second, researchers have typically focused on individual moral characteristics, whereas moral characteristics of the group have been neglected. However, through emergent properties, such as norms and processes, groups help define the type and range of relationships and interactions that are likely or permissible (Rubin, Bukowski & Parker, 2006; Veenstra, Dijkstra & Kreager, in press). Processes of social influence among classmates can be rather persuasive for the behaviour of children (Espelage, Holt & Henkel, 2003; Juvonen & Galvan, 2008). Previous research on class characteristics indicates that class normative beliefs about bullying can help explain bullying-related behaviours over and above individual characteristics (Pozzoli, Gini & Vieno, 2012; Salmivalli & Voeten, 2004).

1.1 The present study

The aim of this study was to obtain new insights into the relative contribution of the four aforementioned moral components of Rest's model to bullying as a group process in primary education. To this end, the participant roles of bully, assistant, defender, outsider and victim were compared with regard to moral sensitivity, represented by sympathy, moral reasoning, moral motivation, and moral personality characteristics at both the individual and class level. We studied these relationships in middle childhood since the amount of time spent with peers and the types of peer-group relationships is quite expansive in this developmental period (Rubin, Bukowski & Laursen, 2011). Moreover, bullying tends to peak during the middle school years (Pellegrini & Bartini, 2000; Rios-Ellis, Bellamy & Shoji, 2000). To determine the unique impact of moral components to bullying, four control variables were taken into account: gender, age, scholastic

ability and socioeconomic status. All have repeatedly been found to be interwoven with children's behaviour in bully situations (for gender see Pozzoli & Gini, 2010; 2012; Seals & Young, 2003; Warden, Cheyne, Christie, Fitzpatrick & Reid, 2003; Warden & Mackinnon, 2003; for age see Espelage, Bosworth & Simon, 2001; Pellegrini & Long, 2002; Scheithauer, Hayer, Petermann & Jugert, 2006; for scholastic ability see Dake, Price & Telljohann, 2003; Jansen et al., 2012; and for socioeconomic status see Dodge, Greenberg & Malone, 2008; Rueden, Gosch, Rajmil, Bisegger & Ravens-Sieberer, 2006; Tippett & Wolke, 2014).

We expected to find lower levels of moral sensitivity, represented by sympathy, for bullies and assistants and higher levels for defenders than for the other participant roles. Previous research shows that sympathy has been negatively correlated with different forms of antisocial behaviour (Björkqvist, Österman & Kaukiainen, 2000; Kaukiainen et al., 1999; Kaukiainen, Björkqvist, Österman & Lagerspetz, 1996; Zahn-Waxler, Cole, Welsh & Fox, 1995) and with bullying in particular (Gini, Pozzoli & Hauser, 2011; Jolliffe & Farrington, 2006). Specifically, Arsenio and Lemerise (2001) and Sutton, Smith and Swettenham (1999) found that bullies may lack the ability to appreciate the emotional consequences of their behaviours on others' feelings, and to sympathize with the feelings of others. In line with this, Caravita, Di Blaso and Salmivalli (2010) concluded that bullies were less able to understand the pain of other children. On the other hand, they found sympathy to be positively associated with defending. As Arsenio and Lemerise (2004) and Hoffman (2000) argued, children's capacity for sympathy is a key contributor to doing the good. Indeed, defending behaviour has been associated with feeling sorry for another person's situation or perspective (Gini, Albiero, Benelli & Altoè, 2007; 2008; Nickerson, Mele & Princiotta, 2008; Stavrinides, Georgiou & Theofanous, 2010). In a review of Van Noorden, Haselager, Cillessen and Bukowski (2015) it was concluded that sympathy showed no association with being victimized and a negative or no association with bystanding.

Our hypotheses regarding associations between bullying and moral reasoning were more open-ended. Two studies within the age range of this study indicate that bullies know just as well as defenders that some reasons to explain behaviour are morally wrong (Gini, Pozzoli & Hauser, 2011; Olthof, 2010). It therefore seems that knowing the good is not sufficient to do the good (e.g., Nucci, 2001). In accordance, Haidt (2001) claims that moral reasoning only serves to justify a moral judgment that has quite different origins. On the other hand, Gini (2006)

found that 8 to 11 year old bullies, assistants and reinforces showed more and victims less moral disengaged reasoning than defenders and outsiders. Perren, Gutzwiller-Helfenfinger, Malti and Hymel (2012) found that bullies between age 12 and 18 use less moral reasoning and more egocentric reasoning than outsiders when presented with hypothetical moral dilemmas. Additionally, they found that victims produced more victim-oriented justifications but fewer moral rules than outsiders. Menesini, Sanchez, Fonzi, Ortega, Costabile and Lo Feudo (2003) found a similar pattern in 14 to 18 year olds when comparing bullies to victims and outsiders.

Bullies and assistants were expected to have deficits in moral motivation compared to the other participant roles. Additionally, defenders were expected to show higher moral motivation than the other participant roles. A recent meta-analysis by Malti and Krettenauer (2013) showed moderate-size relations between antisocial behaviour and emotions following hypothetical moral transgressions, a commonly used measure of moral motivation. Additionally, they found small-size relations between negative anticipated emotions and prosocial behaviour. For behaviour in bullying situations, comparable results were found. Gasser and Keller (2009) report that bullies showed a deficit in moral motivation compared to prosocial children. In the study of Menesini et al. (2003) bullies, as compared to victims and outsiders, showed less negative emotion attributions. On the other hand, defenders attributed more negative emotions to a wrongdoer than victims. In line with this, Sutton et. al. (1999) argue that bullies are willing to initiate intentional aggression that they would otherwise consider unacceptable when their needs conflict with those of others. Contrastingly, Gini (2006) and Menesini and Camodeca (2008) found that guilt was related to defending, but not to bullying and outsider behaviour.

Moral character, represented by inhibitory control and by the personality dimensions agreeableness and conscientiousness, was expected to be negatively associated with bullying, assisting and being a victim and expected to be positively associated with defending. It has been found that bullies and assistants tend to score lower on agreeableness, conscientiousness and inhibitory control compared to other participant roles (Fossati, Borroni & Maffei,2012; Menesini, Camodeca & Nocentini, 2010; Miller, Lynam & Leukefeld, 2003; Tani, Greenman, Schneider & Fregoso, 2003). Additionally, victims score lower on agreeableness and conscientiousness in comparison to other children (Bollmer, Harris & Milich, 2006; De Bolle & Tackett, 2013; Jensen-Campbell & Malcolm, 2007; Tani et al., 2003). On

the other hand, the personality of moral exemplars has repeatedly been found to orient toward conscientiousness and agreeableness (Walker, 1999; Walker & Hennig, 2004). Indeed, in the study of Tani et al. (2003) defenders exhibited high levels of agreeableness in comparison to other participant roles. Generally, children who have good control over their emotions are more likely to exhibit prosocial behaviours (Beauchaine et al., 2013; Carlo, Crockett, Wolff & Beal, 2012; Laible, Carlo, Panfile, Eye & Parker, 2010; Padilla-Walker & Christensen; 2011).

A plausible assumption is that moral components at the class level explain between-class variation in bullying-related behaviours. A greater likelihood of bullying or assistant behaviour in school classes is likely to be negatively associated with collective moral components in the classroom. Conversely, class moral components are expected to be positively related to defending in classrooms. Indeed, Pozzoli, Gini and Vieno (2012) found a positive association between bullying, assisting and reinforcing and class moral disengagement. Also, Gini, Pozzoli and Bussey (2014; 2015) demonstrated that aggression and passive bystanding were more frequent in classes characterized by higher levels of classroom collective moral disengagement. The opposite was true for defending.

2. Method

2.1 Procedure and Participants

Ethical consent for this study was obtained from the Ethical Committee Pedagogical and Educational Sciences from the University of Groningen. Participants were recruited via the personal network of the researchers. First, school principals and teachers were asked for consent. Parental consent letters were then distributed to obtain permission for their children's participation (acceptance rate: 99%). The participants of this study were 1,258 children aged 6 to 13 years (*M* age = 9.10, *SD* = 1.81, 49% females) divided over 54 classrooms attending 11 Dutch primary schools**.** The average number of children per classroom was 24.3 (SD=5.1; range= 14 to 32). The children attended regular education and were predominantly white and of Dutch descent (93.7%).

2.2 Measures

All children participated in an one-on-one interview and filled in an online questionnaire administered by undergraduate students. Children in grade 1 and

those with difficulty concentrating and/or reading were assessed one-on-one, where the researcher read out the online questionnaire. Otherwise, the children were seated in groups of four to ten at computers spaced sufficiently to ensure privacy. The students administering the interview and questionnaire all received extensive training. The children were instructed to provide their own responses to the questions in the questionnaire and interview and were informed that there were no right or wrong answers. Great care was taken to assure that their answers would remain strictly confidential and would not be revealed to anyone. The online questionnaire and the interview took 15-25 minutes each. The interviews were recorded and transcribed afterwards.

2.2.1 Outcome variable

Participant roles in the bullying process. A sociometric questionnaire informed about the perceived participant roles in the bullying process of peers in the classroom, similar to nomination measures used in past research (e.g. Goossens, Olthof & Dekker, 2006). The online questions about the participant roles were preceded by an explanation of the concept of bullying and two questions to verify their understanding of the concept. In the case children provided a wrong answer, the concept of bullying was explained by the research assistant until they understood the concept. Then, the children were then asked whether they were bullied. Depending on the answer to this question, the children were asked to name the classmate(s) that bullied them: "Who starts bullying you?", or to name the classmate(s) that bullied someone else in their classroom: "Who starts bullying?". A list of all their classmates was displayed in random order and they could click on the name(s) of their classmates that fit the description. The children could name an unlimited number of classmates. Children from other classes could also be mentioned using an open question. Other questions concerned joining the bully: "Who joins bullying you?" or: "Who joins bullying?", helping the victim: "Who helps you when you are bullied?" or: "Who helps the victim of the bullying?", and being victimized: "Who gets bullied?" in case the child was not bullied. Three different forms of bullying were distinguished and explained by drawings preceding the questions: (a) physical bullying (i.e., physical peer aggression, such as hitting, kicking or pushing); (b) verbal bullying (i.e., behaviours such as calling names or saying mean or unkind things and (c) relational bullying, a concept that referred mainly to social exclusion (e.g., Perren & Alsaker, 2006). Object-related

bullying (e.g., taking away or breaking other child's belongings) and other specific forms of indirect bullying were left out, since it yielded least responses in the pilot study (Meijer, 2012). Instead, children could specify other forms of bullying in an open question format. Nominations were counted and divided by the number of participating classmates so that the data were comparable across classes. This yielded proportion scores (range 0-1) that indicated by what proportion of classmates each child was named as a bully, assistant, victim, outsider or defender. As such, peer nominations were aggregated across multiple nominators, which enhanced the reliability and validity of the data. The chance of an error occurring due to a single reporter's experience with the child was significantly reduced. Moreover, as insiders, peers can identify characteristics and relationships of children that are considered relevant from the perspective of those who ultimately determine a child's social status and integration in the peer group (Rubin, Bukowski & Parker, 1998). A recommended classification method was used to assign participant roles to the children. The proportion score for bullying was calculated by taking the mean of the proportion scores of the two forms of bullying that the child was most nominated for. Taking the mean of two forms reduces the chance of underestimating the bullying behaviour (Olthof, Goossens, Vermande, Aleva & Van der Meulen, 2011). In a similar manner, proportion scores were calculated for assistant, defender and victim. Children were assigned the role for which they scored .10 or higher and higher than the proportion scores on the other roles. Children that scored lower than .10 on all the roles were assigned the outsider role. Based on by Goossens et al. (2006) we used an absolute instead of a relative criterion. Moreover, we used the lowest criterion of .10 in order to create a homogeneous reference group of outsiders.

2.2.2 Predictors

Moral sensitivity. In this study, moral sensitivity, expressed as sympathy, was assessed with eight statements based on the empathic concern measure of Zhou, Valiente, and Eisenberg (2003) in the online questionnaire. After each statement (e.g. "When I see someone being picked on, I feel sorry for them") children were asked whether the sentence described him/her or not, and if so, how strongly on a scale from 0 to 2: "No, this does not sound like me", "This is sort of like me", and "This is really like me" (α=.81).

Moral reasoning. Moral reasoning was assessed by means of an interview

using a series of six depicted hypothetical transgressions covering three moral domains: fairness (not winning fairly, not keeping word), victimization (verbal bullying) and omission of prosocial duties (refusing to share pencils, refusing to help someone in pain, refusing to stand up for someone) (Jansma, Malti, Opdenakker & Van der Werf, 2017). In the beginning of each scenario, the children were asked why it was right or wrong to transgress the moral rule. Using a validated coding system (Malti, Gasser & Buchmann, 2009) children's' reasons were coded as either moral (i.e., those which refer to moral norms and empathic concern for the victim), non-moral, or other/ unclassifiable. Interrater agreement on the coding was K=.87 for winning fairly, K= .91 for keeping word, K= .88 for sharing pencils, K= .94 for helping someone in pain, K= .88 for verbal bullying and K= .83 for not standing up for someone. The reliability of the scale scores as measured with Cronbach's Alfa was .29 (k=6).

Moral motivation. Moral motivation was assessed using the same interview using six validated moral transgressions. Children were asked how they would feel if they transgressed the moral rule (emotion attribution). By attributing an emotion to a hypothetical wrongdoer, children may indicate the relative importance they attach to moral conformity versus need satisfaction when needs conflict with norms. Following Jansma, Malti, Opdenakker and Van der Werf (2017) anticipated emotions were coded as negative (e.g., bad or half well and half bad) or positive (e.g., happy) emotions. Intercoder reliability of the binary coding of emotions was K=.97 for winning fairly, K= .94 for keeping word, K= .97 for sharing pencils, K= .99 for helping someone in pain, K= .94 for verbal bullying and K= 1.00 for not standing up for someone. The reliability of the scale scores as measured with Cronbach's Alfa was .76 (k=6).

Moral character. In this study, moral character is represented by inhibitory control and the personality traits agreeableness and conscientiousness. Inhibitory control was assessed with an adjusted and translated version of the subscale "Inhibitory Control" of the Early Adolescent Temperament Questionnaire-Revised (Ellis & Rothbart, 1999). Children reported online (e.g. "I am good at self-discipline.") on a scale ranging from 1 (never) to 5 (almost always) (k=11, α =.63). Teachers ratings of the personality dimensions agreeableness and conscientiousness were used. These teacher ratings were derived from the Five-Factor Personality Inventory (FFPI) assessing the Big Five factors of personality (Hendriks, 1997). The five factor scores of the FFPI appeared to be stable and valid in the normal

population in the Dutch cohort study COOL 5-18 (Driessen, Mulder, Ledoux, Roeleveld & Van der Veen, 2009). Additionally, the pupil items of the FFPI were found to be highly related to the teacher ratings on the five personality factors (COOL 5-18, 2008).

Class level moral sensitivity, moral reasoning, moral motivation and moral character. In order to create class level predictors, the mean scores of moral sensitivity, moral reasoning, moral motivation and moral character were calculated for each class and assigned to every child in that class. The scale of the class level variables remained the same as for the individual variables.

2.2.3 Covariates

Age, gender, socioeconomic status, and scholastic ability. Age, gender, socioeconomic status, and scholastic ability will be included as covariates. Socioeconomic status of the children was measured by averaging the level of highest completed education of their father and mother on a scale from 1 to 7 (1=kindergarten, 7=university+). Scholastic ability was measured by taking the average of nationally normed achievement tests on math and reading. Scores range from 1 (lowest score) to 5 (highest score).

2.3 Statistical Analysis

The relations between the moral components and the participant roles was examined with a multilevel multinomial regression analysis using MLwiN (Rasbash, Charlton, Browne, Healy & Cameron, 2005), adjusting for the covariates socioeconomic status, scholastic ability, gender and age. A three-level multilevel analysis was carried out consisting of lower-level observations (i.e. level 1 units of analysis) nested within higher-level observations (i.e. level 2 units of analysis) nested within even higher-level observations (i.e. level 3 units of analysis). In the context of the present study, children were nested within classrooms within schools. Multilevel analysis takes into account this data structure by using a hierarchical linear model that allows for within group variability as well as between group variability (Snijders & Bosker, 2011). A multilevel multinomial logistic model is the most frequently used multilevel model for nominal outcomes and is similar to a logistic regression model except that it allows for the dependent variable to have more than two categories (Rasbash, Steele, Browne & Goldstein, 2009), i.e. bully, assistant, victim, defender, and outsider. With these five possible outcomes, the

model is roughly equivalent to running four binary multilevel logistic regressions comparing outcomes to a reference category, in our case the outsiders. In the multilevel multinomial logistic model, however, all of the logits (i.e. bully to outsider, assistant to outsider, victim to outsider, and defender to outsider) are estimated simultaneously, which enforces the logical relations among the parameters and uses the data more efficiently (Long, 1997).

The predictors of interest were individual- and class-specific and were included in the model as level 1 and level 2 predictors. There were no predictors at level 3; the school level. The starting point of the multilevel multinomial logistic model was the so-called empty model without any predictors. The empty model provided preliminary information about the variance of the dependent variable between classrooms within schools (i.e. level 2) and between schools (i.e. level 3). In the case of a multilevel logistic regression the level 1 variance is fixed at $\pi^2/3=3.29$ and is not estimated. Each subsequent model was compared to the preceding one to evaluate whether the inclusion of additional predictors provided a better fit of the data. In order to take a more or less straight way through the jungle of possible multilevel models, three forward steps after the empty model were distinguished: (1) adding fixed predictors at level 1 (i.e. control variables); fixed effects do not vary across classrooms and can be regarded as the average effect over the whole population of children, (2) adding the fixed explanatory variables of main interest at level 1 (i.e. moral components) to evaluate the unique role in the prediction of the dependent variable while controlling for the variables entered in the previous model step, and (3) adding the fixed effects of the level 2 variables. Having tested for random slopes and included those that were significant ($p<0.05$), the significance of both the fixed and interaction effects was evaluated with the t-test, based on the ratio of parameter estimate to standard error. Insignificant random slopes and interaction effects were removed from the model and this resulted in the final model. For ease of interpretation as well as estimation, the level 1 and level 2 predictors were centred around the grand mean prior to statistical analysis.

3. Results

3.1 Missing data

In order to compute the multilevel multinomial logistic regression model in MLwiN, missing data needed to be deleted in a listwise manner. Unfortunately, not

all teachers filled in the questionnaire about the personality characteristics of the children in their class leading to a reduction of the number of children that could be included in the present study. Furthermore, the remaining data contained some missing values, mainly due to unit non-response or schools not being able to provide background information on parental education or the scholastic ability tests. The 900 children who eventually participated in the current inquiry did not differ from the other children in terms of age, gender, inhibitory control, and moral reasoning ($t(1256)=1.50$; $p=.13$; $\chi^2(1)=.05$; $p=.83$; $t(1256)=-.40$; $p=.69$; $t(1175)=-.83$; $p=.41$). However, non-participating children had a lower mean score on agreeableness, moral emotions, and socioeconomic status than participating children (respectively $t(993)=-2.13$; $p=.03$, $d=.24$; $t(1175)=-2.71$; $p<.01$, $d=.19$; $t(1226)=-2.94$, $p<.01$, $d=.19$) and a higher mean score on scholastic ability and sympathy ($t(1222)=-2.72$, $p<.01$, $d=.17$; $t(1256)=-4.71$, $p<.01$, $d=.29$) and were underrepresented in the outsider participant role ($\chi^2(4)=11.2$; $p=.02$). Nevertheless, these differences were relatively small.

3.2 Descriptive statistics

Table 1 displays the descriptive statistics for the main characteristics of the five different participant roles, i.e. bully, victim, assistant, defender, and outsider. No differences between the participant roles were found concerning scholastic ability, moral reasoning, and moral emotions ($F(4)=1.23$, $p=.39$; $F(4)=0.91$, $p=.46$; $F(4)=.28$, $p=.89$ respectively) and sympathy and moral emotions at the class level ($F(4)=0.93$, $p=.45$; $F(4)=2.08$, $p=.08$). The participant roles differed significantly from one another with regard to gender, age, socioeconomic status, sympathy, inhibitory control, agreeableness and conscientiousness (respectively $\chi^2(4)=102.77$, $p<.01$; $F(4)=37.21$, $p<.01$; $F(4)=3.94$, $p<.01$; $F(4)=3.70$, $p<.01$; $F(4)=4.23$, $p<.01$; $F(4)=36.88$, $p<.01$; $F(4)=13.41$; $p<.01$), and moral reasoning, inhibitory control, agreeableness and conscientiousness at the class level (respectively $F(4)=6.13$, $p<.01$; $F(4)=3.86$; $p<.01$; $F(4)= 7.92$, $p<.01$; $F(4)= 6.61$, $p<.01$). Boys are overrepresented in the bully and assistant group (83.3 and 94.0%) and girls are overrepresented in the victim group (76.9%). Moreover, it appears that assistants score lower on sympathy, inhibitory control, conscientiousness and agreeableness than the other participant roles. Bullies also tend to score lower on agreeableness and conscientiousness. At the class level, the differences between the participant roles seem small.

Table 1

Descriptive characteristics of bullies, assistants, defenders, victims, and outsiders

	Bully	Assistant	Defender	Victim	Outsider
M Sympathy (SD)	2.07 (.50)	1.92 (.46)	2.18 (.43)	2.19 (.53)	2.14 (.47)
M Moral reasoning (SD)	1.83 (.16)	1.81 (.17)	1.79 (.20)	1.78 (.17)	1.81 (.18)
M Moral emotions (SD)	.83 (.24)	.79 (.25)	.81 (.27)	.83 (.27)	.81 (.27)
M Inhibitory control (SD)	3.83 (.49)	3.75 (.47)	3.96 (.50)	4.05 (.43)	3.95 (.47)
M Agreeableness (SD)	2.60(1.16)	2.88 (1.00)	3.56 (.98)	4.06 (.85)	3.77 (.98)
M Conscientiousness (SD)	2.51 (1.18)	2.56 (.96)	3.22 (1.06)	3.35 (1.15)	3.27 (1.09)
M Class Sympathy (SD)	2.13 (.02)	2.11 (.02)	2.15 (.01)	2.11 (.02)	2.14 (.01)
M Class Moral reasoning (SD)	1.81 (.01)	1.81 (.01)	1.78 (.01)	1.81 (.01)	1.81 (.00)
M Class Moral emotions (SD)	.83 (.01)	.80 (.01)	.81 (.01)	.82 (.01)	.81 (.00)
M Class Inhibitory contr (SD)	3.92 (.02)	3.89 (.02)	3.92 (.01)	3.97 (.02)	3.94 (.01)
M Class Agreeableness (SD)	3.45 (.05)	3.41 (.06)	3.55 (.04)	3.67 (.05)	3.66 (.02)
M Class Conscientiousn (SD)	3.08 (.04)	2.93 (.05)	3.19 (.03)	3.13 (.04)	3.19 (.02)
M Scholastic ability (SD)	3.37 (1.10)	3.61 (.93)	3.49 (1.10)	3.67 (1.00)	3.46 (1.11)
M Socioeconomic status (SD)	4.12 (.94)	4.11 (.80)	4.15 (.77)	4.30 (.82)	4.39 (.86)
M Age (SD)	8.86 (1.66)	8.64 (1.49)	7.71 (1.40)	8.25 (1.34)	9.53 (1.77)
% boys	83.3	94.0	41.3	23.1	48.4
N	84	50	126	78	562

3.3 Multilevel multinomial logistic regression: predicting participant roles in the bullying process

First, the multilevel multinomial empty model was estimated to obtain the intra-class correlation coefficients indicating the proportion of variance accounted for at the class level. The coefficients were .02, .16, .16 and .32 for the contrasts comparing bullies, assistants, defenders and victims versus outsiders respectively. At the school level, the intra-class correlation coefficients were all zero. The random effect covariances between the contrasts were positive, indicating that classrooms with a high (low) proportion of bullies tend to have a high (low) proportion of children that are assistant, victim or defender and vice versa.

Table 2 shows the results for the multilevel multinomial models with

predictors at the individual and class level for each of the contrasts comparing the bullies, assistants, defenders and victims to the reference category, i.e. the outsiders. The final model still contains classroom level variation in the prevalence of the participant roles that could not be explained by the predictor variables included in the model. However, the variances did decrease in value in comparison with the child-level model and the empty model. In total, the predictor variables respectively explained 51.0%, 52.9%, 35.3% and 48.4% of the classroom level variation in the prevalence of the categories bully, assistant, defender, and victim in comparison with the outsider category.

Table 2

Multinomial multilevel regression model predicting the participant roles of bully, assistant, defender, and victim versus outsider

	Individual-level predictors		Individual - and class-level predictors**	
Fixed effects	*Estimate*	*S.E.*	*Estimate*	*S.E.*
Intercept - Bully	-2.02	.20	-2.16	.21
Intercept - Assistant	-2.45	.29	-2.77	.31
Intercept - Defender	-2.18	.26	-2.32	.26
Intercept - Victim	-3.00	.36	-2.20	.25
Level 1 (individual) variables				
Gender (girl) - Bully	-1.39*	.33	-1.50*	.34
Gender (girl) - Assistant	-2.57*	.65	-2.87*	.64
Gender (girl) - Defender	.32	.23	.39	.23
Gender (girl) - Victim	1.25*	.31	1.28*	.31
Age - Bully	-.13	.08	-.27*	.09
Age - Assistant	-.12	.14	-.03	.17
Age - Defender	-.59*	.11	-.64*	.12
Age - Victim	-.21	.13	-.34*	.14
Socioeconomic status - Bully	-.27	.16	-.31	.16
Socioeconomic status - Assistant	-.51*	.22	-.45*	.22
Socioeconomic status - Defender	-.03	.15	-.03	.15
Socioeconomic status - Victim	-.14	.17	-.19	.17
Scholastic ability - Bully	.10	.13	.14	.13
Scholastic ability - Assistant	.37*	.18	.38*	.17
Scholastic ability - Defender	-.08	.11	-.08	.11
Scholastic ability - Victim	.23	.13	.23	.13

Sympathy - Bully	.21	.29	.32	.30
Sympathy - Assistant	-.62	.43	-.80	.46
Sympathy- Defender	.11	.25	.89*	.36
Sympathy - Victim	.10	.30	.18	.30
Moral reasoning - Bully	2.60*	.81	2.83*	.85
Moral reasoning - Assistant	2.39*	1.05	2.74*	1.05
Moral reasoning - Defender	.98	.66	1.15	.70
Moral reasoning - Victim	-1.63*	.72	-1.83*	.74
Moral emotions - Bully	-.35	.53	-.42	.58
Moral emotions - Assistant	-.40	.67	-.16	.68
Moral emotions - Defender	1.11	.67	1.05	.70
Moral emotions - Victim	.34	.49	.38	.51
Agreeableness - Bully	-.95*	.13	-1.12*	.15
Agreeableness - Assistant	-.70*	.19	-.86*	.19
Agreeableness - Defender	-.21	.11	-.20	.12
Agreeableness - Victim	.20	.15	.21	.16
Conscientiousness - Bully	-.29*	.13	-.32*	.13
Conscientiousness - Assistant	-.17	.17	-.08	.18
Conscientiousness - Defender	.00	.11	.06	.11
Conscientiousness - Victim	-.08	.13	-.08	.13
Inhibitory control - Bully	-.40	.31	-.39	.31
Inhibitory control - Assistant	.01	.38	-.16	.37
Inhibitory control - Defender	.17	.24	.13	.24
Inhibitory control - Victim	.19	.29	.14	.29
Inhibitory control*Conscientiousness - Bully	-.62*	.25	-.73*	.24
Age*Agreeableness - Assistant	.13	.09	.22*	.09
Sympathy*Agreeableness - Assistant	-.30	.30	-.76*	.29
Sympathy*Inhibitory control - Assistant	.88	.65	1.37*	.59
Age*Moral emotions - Defender	.74*	.31	.83*	.32
Level 2 (class) variables				
Sympathy - Bully			.02	1.07
Sympathy - Assistant			-.17	1.72
Sympathy - Defender			-.16	1.50
Sympathy - Victim			-3.23	1.92
Moral reasoning - Bully			3.04	2.60
Moral reasoning - Assistant			3.13	4.55
Moral reasoning - Defender			-.02	2.74
Moral reasoning - Victim			3.55	4.39

Moral emotions - Bully	5.32*	2.21
Moral emotions - Assistant	-2.46	3.47
Moral emotions - Defender	5.37	2.74
Moral emotions - Victim	4.96	3.49
Agreeableness - Bully	.26	.35
Agreeableness - Assistant	-1.05	.62
Agreeableness - Defender	-.27	.49
Agreeableness - Victim	.28	.60
Conscientiousness - Bully	.05	.40
Conscientiousness - Assistant	-1.18	.71
Conscientiousness - Defender	.92	.61
Conscientiousness - Victim	.20	.74
Inhibitory control - Bully	-2.88*	1.21
Inhibitory control - Assistant	-1.80	1.89
Inhibitory control - Defender	-1.17	1.84
Inhibitory control - Victim	3.61	2.34

Random effects	*Var. Comp.*	*S.E.*	*Var. Comp.*	*S.E.*
School level variance- Bully	.00	.00	-	-
School level variance - Assistant	.00	.00	-	-
School level variance - Defender	.06	.17	-	-
School level variance - Victim	.00	.00	-	-
Class level variance - Bully	.11	16	.08	.01
Class level variance - Assistant	1.09	.47	.71	.34
Class level variance - Defender	1.03	.34	.96	.33
Class level variance - Victim	2.34	.66	1.53	.50
Covariance - Bully, Assistant	.18	.19	.00	.00
Covariance - Bully, Defender	.39	.17	.00	.00
Covariance - Bully, Victim	.15	.23	.00	.00
Covariance - Assistant, Defender	.20	.29	.51	.26
Covariance - Assistant, Victim	.40	.40	.80	.33
Covariance - Defender, Victim	-.12	.34	-.07	.29

* p<.05 **The model did not converge with random effects at the school level, therefore, these effects were removed from the model.

In both models assistants and bullies were predominantly male, whereas defenders were predominantly female, reflecting reality in classrooms (Warder, Cheyne, Christie, Fitzpatrick & Reid, 2003) (respectively t(884)=-4.48; p<.01; t(884)=-

4.41; p<.01; t(884)=4.13; p<.01). In comparison to the outsider group, bullies, defenders and victims were significantly younger in the final model (respectively t(884)=-3.00; p<.01; t(884)=-5.33; p<.01; t(884)=2.43; p<.01). Within the model using child-level predictors only defenders were significantly younger. In both models assistant had a significantly lower socioeconomic status and a significantly higher scholastic ability than the outsider group (respectively t(884)=-2.05; p<.01; t(884)=2.23; p<.01). In the model using child- and class-level predictors, defenders had a significantly higher score on sympathy than outsiders (t(884)=2.47; p<.01). Sympathy was also significantly higher for defenders in comparison to assistants ($\chi^2(1)=7.56$, p<.01). Specifically, the multinomial logit estimate for one unit increase in sympathy is .89 for defenders relative to outsiders given the other variables in the model are held constant. This means that one unit increase in sympathy results in 2.44 (odds ratio = exp(.89)) change in the odds for being a defender relative to an outsider. Thus, the relative risk of being in the defender category increases by 144% when sympathy increases one unit. Also, the probability of being a defender versus an outsider changes with .94 (2.44/1+2.44) with one unit increase in sympathy. For both models bullying and assisting were positively associated with moral reasoning, whereas being victimized was negatively associated with moral reasoning when compared with the outsiders (respectively t(884)=3.33; p<.01; t(884)=2.61; p<.01; t(884)=-2.58; p<.01 for the final model). The corresponding odds ratios (OR) were 7.69, 7.44, and -4.97. Also, a lower moral reasoning ability increased the chance of being a victim compared to bullies, assistants and defenders (respectively $\chi^2(1)=16.65$, p<.01; $\chi^2(1)=12.48$, p<.01; $\chi^2(1)=7.73$, p<.01). No significant differences were found between the participant roles with regard to moral emotions (defender versus bully, assistant and victim, respectively $\chi^2(1)=2.45$, p=.12; $\chi^2(1)=1.50$, p=.22; $\chi^2(1)=.56$, p=.45). Additionally, agreeableness was negatively related to the roles of bully and assistant in comparison to the outsider role (respectively t(884)=7.47; p<.01; OR=-3.04; t(884)=-4.52; p<.01; OR=-2.33). When comparing bullies and assistants with defenders and victims, bullies and assistants were also significantly less agreeable (for defenders respectively $\chi^2(1)=21.56$, p<.01; $\chi^2(1)=8.53$, p<.01, for victims respectively $\chi^2(1)=37.74$, p<.01; $\chi^2(1)=18.80$, p<.01). Bullies had a significantly lower score on conscientiousness than the outsider group (t(884)=2.46; p<.01; OR=-.87).

Five interesting interaction effects were found. Visualizations of these interaction effects can be found in Appendix E. For both models children their

chance to be an outsider in comparison to a bully increased when inhibitory control was high. This relation was strengthened when conscientiousness was high (t(884)=-3.04; p<.01; OR=0.48). Specifically, the multinomial logit estimate for someone who scored 1SD below the mean of conscientiousness and inhibitory control was -2.18 ((-2.16+(-.32*-1)+(-.39*-1)+(-.73*-1*-1)), corresponding OR=0.11); the multinomial logit estimate for someone who scored 1SD below the mean of conscientiousness and 1SD above the mean of inhibitory control was -1.50 (OR=0.22); the multinomial logit estimate for someone who scored 1SD above the mean of conscientiousness and 1SD below the mean of inhibitory control was -1.36 (OR=0.26); and the multinomial logit estimate for someone who scored 1SD above the mean of conscientiousness and inhibitory control was -3.60 (OR=0.03). Second, when children scored higher on agreeableness they were less often assistants than outsiders. This negative relation was strengthened when children were younger (t(884)=2.44; p<.01; OR=1.25). Contrarily, the negative relation between agreeableness and assisting ceased when children scored lower on sympathy. The relationship was strengthened when children scored higher on sympathy (t(884)=-2.62; p<.01; OR=0.47). Sympathy also played a role in the (negative) relationship between assisting and inhibitory control. When children scored higher on sympathy, children were more likely to be an assistant than outsider when scoring higher on inhibitory control. However, when children scored lower on sympathy, children were less likely to be an assistant versus outsider when scoring higher on inhibitory control (t(884)=2.32; p<.01; OR=3.94). Last, children were more likely to be a defender than an outsider when their moral emotions score increased. This positive relation was strengthened when children were older (t(884)=2.59; p<.01; OR=2.29). Appendix E also provides multinomial logit estimates and odds ratio's corresponding to 1SD above and below the concerning variables for the four aforementioned interaction effects.

Most variables at the class level did not show significant differences between the bully roles. However, class moral emotions increased the chance of being a bully versus being an outsider (t(884)=2.41; p<.01; OR=204.38) and being a victim ($\chi^2(1)=6.93$, p<.01). A negative association appeared between class inhibitory control and bullying versus outsider behaviour (t(884)=-2.38; p<.01; OR=0.06). Assisting was accompanied by a significantly lower score on class conscientiousness than defending ($\chi^2(1)=6.63$, p<.01).

4. Discussion

This study related different participant roles in the bullying process to four components of moral functioning of children age 6 to 12. The components were moral sensitivity, represented by sympathy, moral reasoning, moral motivation and moral character, represented by agreeableness, conscientiousness and inhibitory control. The findings of this study contributed to the existing literature relating moral functioning to bullying by simultaneously linking a broad pallet of moral components to children's behaviour in bullying situations at both the individual and class level.

A relation was expected between the four moral components and the different participant roles involved in bullying: bully, assistant, defender, victim, and outsider. The results provided some support for this proposition. Specifically, sympathy increased the chance of being a defender versus an outsider in the model using child- and class-level predictors. This indicates that especially defenders hold the ability to appreciate the emotional consequences of their behaviours on others' feelings. This ability, as Arsenio and Lemerise (2004) and Hoffman (2000) argued, probably leads children to doing the good in bullying situations. Additionally, as expected, a negative relation between bullying and assisting and sympathy was apparent in the descriptive statistics. This confirms that bullies and assistants are less inclined to let the feelings of others guide their behaviour (e.g. Arsenio & Lemerise, 2001; Sutton, Smith & Swettenham, 1999), even though these relations did not reach significance in the multinomial logistic multilevel analysis. However, an interaction effect was found between sympathy and inhibitory control when comparing assistants to outsiders, indicating that higher sympathy and higher inhibitory control increases the chance that a child is an assistant instead of an outsider. This might indicate that there is a group of assistants with high social skills using them for their own instead of others' profit. This is in line with the Machiavellian intelligence hypothesis, stating that social skills can facilitate not only cooperation, but also competing with others, enabling individuals to gain advantageous positions or manipulate others in order to realize own goals (Björkqvist, Österman & Kaukiainen, 2000; Davis & Stone, 2003). Another explanation for this unexpected interaction effect might be that it is a spurious effect due to multiple comparisons. Confirming the findings of the review of Van

Noorden, Haselager, Cillessen and Bukowski (2015), victims did not significantly differ in sympathy from the outsider group.

Moral reasoning was positively related to bullying and assisting and negatively related to being victimized when compared to the outsider role. In earlier studies, bullies appeared to have a similar level of moral reasoning as defenders (Gini, Pozzoli & Hauser, 2011; Olthof, 2010). However, the results of this study show that moral reasoning even increases belonging to the bully or assisting versus outsider role. This is in accordance with the claim that knowing the good is not sufficient to do the good. Moreover, as the work of Batson, Kobrynowincz, Dinnerstein, Kampf and Wilson (1997) shows, children's' moral reasoning might reflect a desire to appear moral rather than a deeply felt moral commitment. The results also confirm Menesini et al. (2003) and Perren et al. (2012) showing that victims showed less moral reasoning compared to outsiders.

Bullies and assistants were hypothesised to have deficits in moral motivation, as indicated by a lack of negative emotions following moral transgressions. Defenders were expected to show higher moral motivation than the other participant roles.. These hypotheses could not be confirmed by the multinomial multilevel analysis, but the descriptive results confirm the direction of these relations. This suggests, as Sutton et. al. (1999) argue, that bullies as well as assistants are willing to initiate intentional aggression that they would otherwise consider unacceptable when their needs conflict with those of others. For defenders, this was especially true when they were older because we found a significant interaction effect between age and attributed emotions for the comparison between defenders and outsiders. Apparently, anticipated emotions are less important for defending when children were younger. Interestingly, class moral emotions were significantly positively associated with bullying in comparison to outsider behaviour. A plausible explanation for this association may be attributed to the nomination procedure used to determine the participant roles in the bully process. Observant peers might be more prone to report bullying than unobservant peers. Thus, in a classroom with higher levels of negative anticipated emotions after a hypothetical wrongdoing, children might report more bullying-related behaviour than children in classrooms with lower levels of these emotions.

The moral personality characteristics agreeableness, conscientiousness and inhibitory control were expected to be negatively associated with bullying, assisting and being a victim and expected to be positively associated with defending. The

results of the current study indicated that agreeableness was negatively related to the roles of bully and assistant in comparison to the defender and outsider role. Also, the negative relation between assisting and agreeableness became stronger with higher levels of sympathy. This indicates that the personality characteristic agreeableness is an important factor in bullying-related behaviour (e.g. Fossati et al. 2012; Menesini et al., 2010; Miller, et al. 2003; Tani et al., 2003). Further, a negative association appeared between conscientiousness and bullying versus outsider behaviour. Previous research already showed that children with lower self-discipline, orderliness and goal pursuit are more likely to exhibit bully behaviours (e.g. Fossati et al. 2012; Menesini et al., 2010; Miller, et al. 2003; Tani et al., 2003). This negative relation was strengthened when children had lower inhibitory control. Apart from this interaction effect, no associations were found between inhibitory control and the participant roles at the individual-level. Strikingly, however, the results of this study show a relation between inhibitory control and behaviour in bullying situations at the class level. This again indicates that groups help define the type and range of relationships and interactions that are likely or permissible (e.g. Rubin, Bukowski & Parker, 2006) and that class moral characteristics can help explain bullying-related behaviours over and above individual characteristics (Gini et al., 2014; 2015; Pozzoli, Gini & Vieno, 2012). In our case, the collective ability to suppress interferences from the environment is associated with fewer bullies in class. This highlights the importance of class characteristics for explaining bullying-related behaviours over and above individual characteristics (e.g. Pozzoli, Gini & Vieno, 2012; Salmivalli & Voeten, 2004).

Despite its novel focus on relations between dimensions of moral functioning and bullying roles, this study has several limitations. First, the sociometric measures of bullying involved questions of a relational nature and therefore these measurements were interdependent. The use of peer nominations might have been disadvantageous since behavioural reputations sometimes consolidate. Thus, even though a child's behaviour may have changed over the year, their reputation for this behaviour persists with peers. Furthermore, a child's judgment of a peer might be influenced by their own abilities and behaviour. However, besides these two downsides, peer nominations mainly have advantages. The chance of an error occurring due to a single reporter's experience with the child

reduces significantly and peers can identify characteristics and relationships of children that are considered relevant from their perspective (Rubin, Bukowski & Parker, 2006). Another limitation is related to the construction of the participant roles in the bullying process. Specifically, the participant roles were established without distinguishing between different types of bullying. This might have affected the results since previous studies suggest that moral components might be differentially related to verbal and physical forms of bullying versus relational forms of bullying. For example, sympathy related constructs seem to be more related to relational than verbal and physical forms of bullying (Kaukiainen et al., 1996; 1999). Further, the assessment of conscientiousness and agreeableness exclusively relied on teacher-reports. Also, Cronbach's Alpha was relatively low for the scale scores of moral reasoning. Therefore, the results with regard to this measure must be interpreted with care. Future studies could use more advanced measures of both personality characteristics from both the teacher's and child's perspective.

We are also aware of the fact that the data we lost was not missing at random. The children in the sample that was analysed scored higher on sympathy, agreeableness and socioeconomic status than the non-participating children. It could be argued that this selective sample led to an underestimation of the association between moral components and behaviour in bullying situations. Another pattern in the data was related to gender. Whereas victims were predominantly female, assistants and bullies were predominantly male. Although this gender imbalance reflects reality in the children's classrooms, it does raise difficulties of interpretation for the comparison between the role groups (Warder, Cheyne, Christie, Fitzpatrick & Reid, 2003). For example, in the study of Warden and Mackinnon (2003) it was prosocial girls and not prosocial boys who demonstrated greater sympathy-related awareness.

Third, given that the data from this study were only collected at one time point, it only allows for cross-sectional analysis. To consider developmental change, further research should use longitudinal designs that deepen the understanding of how moral precursors influence the participation in the process of bullying.

Despite its limitations, the current study has contributed to the existing literature in several ways. First, this study measured several moral components by following the theoretical framework of Rest and using a broader representation of

measures of moral functioning than most studies do. Interestingly, all the moral components showed specific relations with the chance of taking a particular participant role in the bullying process. This supports an integrated analysis of several moral components for a deeper understanding of the moral precursors of bullying behaviour. Furthermore, it showed the importance of distinguishing between the different participant roles of bully, assistant, defender, victim and outsider. Also, the current inquiry demonstrated that class moral characteristics can help explain bullying-related behaviours over and above individual characteristics. The results of the present study show that the personality characteristic agreeableness is most related to bullying and assisting behaviour when compared with defending and outsider behaviour. Stimulating agreeableness in middle childhood therefore seems a good place to start in order to prevent bullying at school. Moreover, lower conscientiousness was associated with greater chances of bullying (and even more so in combination with lower inhibitory control), and higher sympathy and more negative anticipated emotions levels (and even more so when children were older) were positively associated with defending, in comparison to outsider behaviour. Interestingly, bullies and assistants seem to have higher levels of moral reasoning than outsiders. This indicates that interventions targeting bullying behaviour might not benefit from training moral reasoning skills. As Reiman and Dotger (2008) conclude, telling students what they ought to do is ineffective for promoting prosocial dispositions in children. Additionally, the present study showed that negative anticipated emotions following a moral transgression and low inhibitory control at the class level predicted high ratings of bullying. This suggests that targeting all children in class might be essential for successful interventions. Specifically, intervention programs aimed at affecting bullying might profit from enhancing all children's inhibitory control. Also, these interventions might benefit from either clarifying the consequences of actions to all children to activate their negative emotions or actually increasing the consequences of negative actions in the bullying context. Accordingly, this study has gone some way towards enhancing our understanding of the precursors of children's behaviour in bully situations. Further advancements in the understanding of how individual and class moral components affects bullying-related behaviour may inform educational researchers and practitioners alike about the promotion of prosocial behaviour and interventions concerning antisocial behaviour.

Chapter 5 Promoting Agreeableness in Middle Childhood: Effects of a Class-Based Intervention

Abstract

The aim of this paper was to examine how education might be able to influence children's behavioural tendencies towards agreeableness in order to target bullying-related behaviour in middle childhood. Specifically, this paper investigates the effects of a class-based intervention program promoting agreeableness on bullying-related behaviour and pro- and antisocial behaviour over the course of two school years. After two years, 42 children were not involved in the intervention (52.0% males; M age 9.10), 51 children received the intervention the first year (49.0% males; M age 8.99), 69 children received the intervention the second year (52.0% males; M age 8.50), and 42 children received the intervention during both years (52.0% males; M age 8.91). Promising effects of the class-based intervention were found for all conditions fostering agreeableness and reducing bullying-related behaviour and antisocial behaviour. Positive effects were especially visible in the extended intervention condition. Moreover, a number of gender and age specific effects were found. Overall, findings suggest that promoting agreeableness may serve to counteract antisocial behaviour. We discuss the findings in relation to current educational practice and provide recommendations for future research into success-promoting processes of anti-bully programs.

Note. This chapter is based on Jansma, D.J., Opdenakker, M.C.J.L. & Van der Werf, M.P.C. (2017). Promoting Agreeableness in Middle Childhood: Effects of a Class-Based Intervention. *Manuscript submitted for publication.*

1. Introduction

Antisocial behaviour in education is an extensive problem. Children who exhibit antisocial behaviour are a great burden for teachers, victims, classroom climate, and society as a whole (Hawker & Boulton, 2000; Nansel et al., 2004; Nishina & Juvonen, 2005; Salmivalli & Isaacs, 2005; Soepboer, Veenstra & Verhulst, 2006). Therefore, there is an increasing awareness of the need for education to target antisocial behaviour and encourage the acquisition of prosocial values and behaviour (Cuevas, 2011; Rupp & Veugelers, 2003). In educational practice, this awareness is translated into numerous educational intervention programs aimed at affecting bullying-related behaviour, social emotional learning and moral development (Reiman & Dotger, 2008; Sklad, Diekstra, Ritter, Ben & Gravesteijn, 2012). Unfortunately, studies have shown inconsistent results concerning the effectiveness of these programs. Programs directed at social emotional learning show a considerable variety in efficacy (Weare & Nind, 2011). Also, the majority of studies considering the effectiveness of interventions targeting bullying show nonsignificant outcomes, some negative, and only a few show beneficial outcomes (Merrel, Gueldner, Ross & Isava, 2008; NJI, 2015; Smith, Schneider, Smith & Ananiadou, 2004). Moreover, these beneficial outcomes are consistently distributed across types of assessment, variables and interventions (Merrel et al., 2008). Smith et al. (2004) conclude that anti-bully interventions can succeed, but "not enough is known to indicate exactly how and when" (p. 558). In line with this, Gravemeijer and Kirschner (2007) argue that research aimed at education innovation should not only focus on evidence of effectiveness (What works?) but on understanding the processes explaining the effectiveness (How does it work?).

Therefore, in order to develop effective interventions aimed at affecting anti- and prosocial behaviour and social emotional learning, researchers started to identify specific success-promoting factors of these programs. The current study will try to contribute to this line of research by looking into behavioural characteristics underlying anti- and prosocial behaviour with the aim to reduce bullying-related behaviour in primary education. Bullying is a specific type of antisocial behaviour with the intention to hurt. It is characterized by a repetition of negative actions against a peer and occurs in a context of an imbalance of power, with a more powerful person or group attacking a less powerful one (Olweus, 1993). Most studies understand bullying as immoral action tendency, i.e. behaviour

that is intended to harm others (Gasser, Gutzwiller-Helfenfinger, Latzko & Malti, 2013). Bullying is an especially interesting immoral behaviour as the perpetrator decides intentionally to harm a person in an inferior position which is against moral standards (Vollmeyer, Jenderek &, Tahmine, 2013). Also, both bullying and defending the bully are prime examples of morally relevant behaviour in middle childhood, because of its direct effect on the welfare of victims (Turiel, 1983; 1998). Second, different facets of moral functioning have been found to be important to understand individual differences in engagement in bullying situations (Gini, 2006; Hymel, Rocke-Henderson, Bonanno, 2005; Perren & Gutzwiller-Helfenfinger, 2012). Thus, moral functioning might serve as a central process underlying children's bullying behaviour. Therefore, the present study will focus on moral functioning as a factor being conceptually intertwined with bullying (Cuevas, 2011; Gasser & Keller, 2009) and defending behaviour (Caravita, Gini & Pozzoli, 2012; Menesini & Camodeca, 2008; Pozzoli & Gini, 2010).

Researchers have recently begun to address the role of moral factors in the process of bullying-related behaviour (Arsenio & Lemerise, 2001; 2004; Guerra, Nucci & Huesmann, 1994). The bulk of research studying the relation between moral functioning and behaviour mainly focused on one aspect of moral functioning, namely moral reasoning (Jordan, 2007). However, the relation between moral reasoning and moral behaviour may be quite weak (Blasi, 1980; Thoma & Rest, 1999). Based on a review of psychological research, Rest (1983; 1986) argued that explanations of moral behaviour must not only target moral reasoning but also the ability to interpret correctly what is happening (moral sensitivity), the motivation to behave in a moral fashion (moral motivation), and the ability to persist in a moral task in the face of obstacles (moral character). Thus, Rest's Four Component Model, a widely used framework to assess the psychological antecedents of moral behaviour, postulates four components contributing to moral behaviour: moral sensitivity, moral reasoning, moral motivation and moral character (Myyry, Juujärvi, Pesso, 2010). Jansma et al. (2016) were the first to simultaneously relate all four components to bullying-related behaviour in middle childhood. Their results show that low scores on aspects of moral character, such as agreeableness, inhibitory control and conscientiousness, were most associated with bullying-related behaviour, followed by moral sensitivity.

When following these results, influencing moral character might be a success-promoting factor of interventions targeting bullying-related behaviour in

middle childhood. In children, moral character is expressed by the personality dimensions agreeableness and conscientiousness and by inhibitory control, i.e. behavioural and cognitive suppression of interferences from the environment (Ahadi & Rothbart, 1994; Graziano & Eisenberg, 1997). In the study of Jansma et al. (2016) it was agreeableness, one of the aspects of moral character, that was most associated with bullying-related behaviour. This association transcended gender and age. Agreeableness is one of the big five personality characteristics, the most commonly used five broad domains or dimensions used to describe human personality (Mroczek & Little, 2006). In the meta-analytic review of Miller and Lynam (2001) on the relation between personality and antisocial behaviour, agreeableness showed the highest correlations (greater than .25) with antisocial behaviour of all personality characteristics. Other studies agree that disagreeableness is the strongest and most consistent personality predictor of negative relationship outcome (Shiner, Masten & Tellegen, 2002). This is no wonder since agreeableness reflects individual differences in tendencies towards prosocial behaviour and social harmony (Graziano & Eisenberg, 1997; Tobin, Graziano, Vanman & Tassinary, 2000) and manifests itself in individual behavioural characteristics that are perceived as warm, cooperative, considerate, sympathetic, kind, and generous (Thompson, 2008). The low end of agreeableness includes trait descriptors such as selfish, aggressive, rude, spiteful, teases others, stubborn, bossy, cynical, critical and manipulative. High-disagreeable individuals may escalate negative affect during conflict whereas high-agreeable people are better able to regulate emotions during interpersonal conflicts (Jensen-Campbell & Graziano, 2001).

Within a classroom setting, agreeableness is related to the preferred behavioural pattern of children (Sneed, 2002). High levels of agreeableness have been found to predict positive school outcomes as well as social relations with classmates (e.g., Hampson, Goldberg, Vogt & Dubanoski, 2007; Jensen-Campbell, Gleason, Adams & Malcolm, 2003). On the other side, low agreeableness is considered a risk factor for the development of adjustment problems (Tackett, 2006). Interestingly, agreeableness is assumed to be the most malleable of the personality dimensions and most susceptible to change in light of environmental inputs (Bergeman et al., 1993, Graziano & Eisenberg, 1997). Therefore, the aim of this paper was to examine how education, one of the most important contextual factors in childhood, might be able to influence children's behavioural tendencies

towards agreeableness in order to target bullying-related behaviour. Specifically, this paper investigates the effects of a class-based intervention program promoting agreeableness in middle childhood on bullying-related behaviour, and anti- and prosocial behaviour over the course of two school years.

1.1 Theoretical background of the intervention program

The theoretical rationale underlying our intervention is based on two theoretical accounts of agreeableness, namely interpersonal theory and behavioural theory. According to interpersonal theory, agreeableness is connected to motives for maintaining positive interpersonal relations. In short, agreeableness might allow individuals to minimize the negative impact of conflicts and negotiate outcomes that capitalize on the advantages of group living (Shiner & Masten, 2008). Indeed, agreeableness is manifested in cooperation and friendliness (Mroczek & Little, 2006). According to the behavioural account, agreeableness may emerge developmentally from temperamental self-regulative systems (e.g. McAdams & Adler, 2006; Rothbart & Bates, 1998; Rothbart, Derryberry & Posner, 1994). In line with this, it has been proposed that agreeableness taps into self-control in interpersonal relationships (Tobin et al., 2000). Ahadi and Rothbart (1994) suggest that agreeableness is linked to temperamental bases of effortful control, specifically the regulation of anger. Also, Graziano and Eisenberg (1997) argued that agreeableness is likely to have its origins in the self-regulation of negative emotions. Recent studies provide good evidence for this claim. Disagreeable children may be those whose strong feelings of anger and frustration are not tempered by good self-control (Caspi, Harrington, Milne, Amell, Theodore & Moffit, 2003; Shiner, Caspi, 2002). Additionally, regulated positive emotionality and sociability are likely precursors of later prosocial behaviour (Shiner, 2006). Generally, children who have good control over their emotions are more likely to exhibit prosocial behaviours (Arsenio & Lemerise, 2004; Eisenberg, Wentzel & Harris, 1998). Therefore, the intervention program will focus on the promotion of self-regulation of (negative) emotions and on the promotion of positive interpersonal relations.

Additionally, within research on personality broad higher order traits like agreeableness explain covariation among lower order traits (e.g. prosocial tendencies), and these lower order traits explain covariation among specific behavioural descriptors (e.g., helping) (Halverson et al., 2003; Goldberg, 2001; Putnam, Ellis & Rothbart, 2001). In childhood, agreeableness includes the lower

order traits of prosocial tendencies, gentleness, wilfulness, modesty, integrity, patience, and trust, and their counterparts (Ashton, Lee & De Vries, 2014; Barbaranelli, Caprara, Rabasca & Pastorelli, 2003; Haverson et al. 2003; Mroczek & Little, 2006; Peabody & De Raad, 2002). Since the lower order traits appear to be closer to actual behaviour, targeting the lower order traits of agreeableness seems an efficient way to increase agreeable behaviour and decrease disagreeable behaviour. A meta-analytic review by Jones, Miller and Lynam (2011) showed that all the lower order traits that construct agreeableness are associated with antisocial behaviour. Therefore, the intervention program will focus on the promotion of self-regulation of (negative) emotions and on the promotion of positive interpersonal relations targeting the lower order traits of agreeableness.

1.1.1 Intervention program components

The intervention program was composed of a training for teachers and fourteen one-hour lessons for the pupils taught by an external teacher during the first year. During these lessons, the teacher of the class was also involved. Moreover, the teacher and pupils were given an assignment every week in line with the theme of that week. Each lesson drew upon the previous lesson and the corresponding assignment. During the second year both the lessons and assignments were provided by the teachers of the class in case these teachers engaged in the program the previous year. The teacher training comprised of three training days a year and stressed four ways of reinforcing a safe environment: establishing a group identity, making the pupils responsible, responding to disagreeable behaviour, and responding to agreeable behaviour (e.g. Caprara et al., 2014; Korpershoek, Harms, De Boer, Van Kuijk & Doolaard, 2016; Olweus, 1993). Creating a safe environment is the first condition for a successful implementation of a program directed at behavioural change within pupils (Korpershoek et al., 2016). In addition, a safe environment also stimulates agreeable tendencies like trust and integrity. Therefore, the first three lessons were directed at getting to know each other (better) and creating classroom rules. According to the contact hypothesis (Allport, 1954) contact is one of the best ways to improve relations among groups. When children get to know each other better, they are more inclined to take the perspective of their classmates and this is important for developing agreeableness (Gutzwiller-Helfenfinger, 2008). Rules, on the other hand, create transparency, guidance, and safety. By establishing rules children learn how to take into account

the feelings and wishes of others making it easier to trust their classmates; both characteristics of agreeableness (Jones & Jones, 2010; Korpershoek et al., 2016).

After the first three lessons, the program mainly focussed on the promotion of self-regulation of negative emotions, since agreeableness and disagreeableness are strongly linked to self-control in interpersonal situations. In the teacher training we also focused on dealing with (negative) emotions within the classroom setting. Teaching children to recognize emotions in themselves and others and to express them in a mild way when experiencing negative emotions helps them to better control their emotions (Urbain & Kendall, 1980; Greenberg, Kusche, Cook & Quamma, 1995; Warden & Mackinnon, 2003). Next to these important facets of the program, we also payed substantial attention to cooperation and trust in both the lessons and the teacher training. Interpersonal theory namely stresses the importance of agreeableness in the light of cooperation and trust. Cooperative methods put children's patience to the test and provide encounters with disappointment and frustration. When regulation these negative emotions, self-control comes into play (Eisenberg, Spinrad & Eggum, 2010; Singer, Golinkoff & Hirsh-Pasek, 2006). Also, as said before, trust is a lower order trait of agreeableness. Thus, both the lessons and teacher training had four overarching themes directed at promoting agreeableness: getting to know each other, rules, emotions and cooperation and trust (Eisenberg et al., 2010; Jones & Jones, 2010; Warden & Mackinnon, 2003).

1.2 The present study

The aim of this paper was to examine how education might be able to influence children's behavioural tendencies towards agreeableness in order to target bullying-related behaviour in children age 6 to 12. Specifically, this paper investigates the effects of a class-based intervention program promoting agreeableness on bullying-related behaviour and prosocial and antisocial behaviour. Because the seeds of anti- and prosocial behaviour most probably emerge in childhood (Eisenberg et al., 2002), knowledge about these years is most essential to understand and intervene in anti- and prosocial behaviour in education. Moreover, children aged 7 to 9 have rarely been investigated in research on bullying and therefore form an interesting target group (Gasser & Keller, 2009). The program effects were examined by comparing children in a control condition with three intervention conditions at the beginning and at the end of two school years: (a) the

intervention condition in which the intervention was implemented during the first schoolyear, (b) the delayed intervention condition in which the intervention was implemented during the second schoolyear, and (c) the extended intervention condition in which the intervention was implemented during both school years.

We hypothesized that the intervention program would be able to increase agreeableness in all conditions. Moreover, we assumed that this increase in agreeableness might reduce antisocial behaviour (McCrae & John, 1992; Shiner, 2006) and bullying-related behaviour (Cook, Williams, Guerra, Kim & Sadek, 2010), and increase prosocial behaviour (Graziano & Eisenberg, 1997; Jensen-Campbell, Graziano & Hair, 1996; Sneed, 2002). With regard to bullying-related behaviour we expected a decrease in the prevalence of bullying, assisting, outsider behaviour, and victimization and an increase in the prevalence of defending (Tani, Greenman, Schneider & Fregoso, 2003). The biggest effects in the aforementioned outcomes were expected within the extended intervention condition. Within the delayed intervention condition we expected to see effects only during the second schoolyear. For the intervention condition we expected effects in the first schoolyear with a continuation of these effects in the second schoolyear.

As previous studies have shown that gender and age are important predictors of bullying-related behaviour, anti- and prosocial behaviour as well as agreeableness (e.g. Eisenberg, Spinrad & Knafo, 2015; Salmivalli & Voeten, 2004; Srivastava et al., 2003; Whitney & Smith, 1993), they were included as covariates in our analyses. In this way, their effects were controlled for and their potential interactions with the intervention could be examined. Scholastic ability and socioeconomic status were also included as covariates (e.g. Downey, Mountstephens, Lloyd, Hansen & Stough, 2008; Donnellan, Trzesniewski, Robins, Moffit, Caspi, 2005; Tippett & Wolke, 2014).

2. Method

2.1 Design and sample

The present study was part of a longitudinal research project examining functioning and prosocial and antisocial behaviour in middle childhood. The project was funded by a NWO (Netherlands Organisation for Scientific Research) Research Talent Grant and by the school participating in the current inquiry. The study makes use of the data collected at a public school located in the Northern part

of the Netherlands recruited via the personal network of the researcher. Ethical approval for this study was obtained from the Ethical Committee Pedagogical and Educational Sciences from the University of Groningen. First, the school principal and teachers were asked for consent. Hereafter, parental consent letters were distributed to obtain permission for their children's participation (acceptance rate: 99%). Data collection took place at the beginning and at the end of two school years: in September/October 2014, May 2015, and in September/October 2015 and March/April 2016. During the first school year grades 1 to 6 were offered the class-based intervention and within the same school, the parallel grades 1 to 6 were the control condition. The lessons of the program were taught by an external teacher. During the second academic year the teachers that were involved in the class-based intervention the previous year taught the intervention themselves. The external teacher also taught the lessons to two classes. After two years, four conditions could be distinguished: children that were not involved in the intervention for two years, i.e. the control condition (n=50; 48.0% males; M age 8.68); children that were involved in the intervention during the first year, but not the second year, i.e. the intervention condition (n=75; 56.0% males; M age 8.34); children that were involved in the intervention during the second year, but not the first year, i.e. the delayed intervention condition (n= 74; 51.0% males; M age 8.44); children that were involved in the intervention for two years, i.e. the extended intervention condition (n=47; 49.0% males; M age 8.85). The average class size was 22.67 (SD=5.12). Both the control and intervention conditions comprised of predominantly white pupils of Dutch descent (97.5%).

2.2 Procedure and intervention

During four measurement occasions, the children filled in an online questionnaire. All questions and measures were formulated and posed in Dutch. The questionnaire was administered by undergraduate students who received extensive training. At all times, the trained test administrator was available for help. Children in grade 1 and those with difficulty concentrating and/or reading were assessed one-on-one, where the researcher read out the questionnaire. Otherwise, the children were seated in groups of four to ten at computers spaced sufficiently to ensure privacy. The online questionnaire was preceded by a general introduction and instructions regarding the measure. The children were instructed to provide their own responses to the questions and were informed that there were no right or

wrong answers. Great care was taken to assure children that their answers would remain strictly confidential and would not be revealed to anyone else and accordingly. Children filled out the online questionnaires during regular school hours.

In between the measurement occasions, the class-based intervention was carried out. The program takes on competency-based approach to teaching, fostering learning through the creation of opportunities for practicing rather than by means of discourse i.e. learning by doing, which has a higher probability of translating into daily actions (Bustamante & Chaux, 2014). The intervention consisted of three training days a year for the teachers and fourteen lessons. The implementation of the program differed between grades 1 and 2, grades 3 and 4, and grades 5 and 6 in order to create an developmentally appropriate program (Jones & Jones, 2010; Malti, Chaparro, Zuffianó & Colasante, 2016). Individual differences between children were also taken into account by varying the types of work during the lessons that were part of the program (e.g. Hoogeveen & Winkels, 2008). In the attachment two example lessons are presented. The first two lessons were devoted to getting to know each other better. The goal was to get the children and the teacher more acquainted. They learned the meaning of each other's first and last name, where everyone lived and what their families looked like. Moreover, the children and teacher learned to name characteristics of one another. Rules were the second theme of the intervention. In the lessons the pupils individually and collectively came up with rules they considered essential for their class (e.g. Thornberg, 2008). The rules were reinforced and concretized with the help of stories, group discussions, cooperative partnerships, physical activities, and movies. The three lessons on trust and cooperation provided the children with cooperative play and cooperative assignments. Moreover, we offered them exercises in trust. During the lessons and the assignment the children we asked to evaluate their own and others behaviour during the exercises. Also, we made use of five essential elements the successful incorporation of cooperative learning in the classroom: positive interdependence, individual and group accountability, promotive face-to-face interaction, teaching the children the required interpersonal and small group skills and group evaluation (Johnson & Johnson, 1999). In order to amplify children's self-control over emotions we first discussed what emotions are and which basic emotions exist. The basic emotions of happiness, sadness, anger and fear were extensively covered during the lessons (Ekman, 1992). The children

discussed whether they experience certain emotions and how they deal with them. In this way, they learned that classmates experience similar emotions, but express them in different ways. Children learned to recognize emotions in themselves and others and learned to express them in a mild way when experiencing negative emotions and conflict (Greenberg et al., 1995; Urbain & Kendall, 1980; Warden & Mackinnon, 2003). Music fragments, drawing, making a group collage, games, stories, singing, practice and multiple role plays were used in the lessons.

2.3 Variables and instrumentation

Agreeableness. Teachers rated each child on a five point scale from 'not agreeable' (bossy, quarrels) to 'agreeable' (kind, mild). These ratings were derived from the Five-Factor Personality Inventory (FFPI) that appeared to be stable and valid in the normal population (Hendriks, 1997). Teacher ratings on the five personality factors are highly related to the pupil items of the FFPI (Driessen, Mulder, Ledoux, Roeleveld & Van der Veen, 2009). For the third and fourth occasion we added four items assessing agreeableness: 'Accepts people as they are', 'Wants to be in charge', 'Respects the opinion of others', 'Imposes his/her will on others'. The item-total correlation with the item we used for the first two occasions was relatively high (T4: r=.70, and T3: r=.71).

Antisocial and prosocial Behaviour. Peer ratings of antisocial ('Who quarrels a lot?' and 'Who says and does mean things?') and prosocial behaviour ('Who helps other children?' and 'Who says and does nice things?') were obtained using an online questionnaire. The children were given a roster with the names of their classmates and they could select as many, or as few, classmates as they wanted. This sociometric method of assessing behaviour has been used in other studies that established its validity (prosocial behaviour: Carlo, Koller, Eisenberg, Da Silva & Frohlich, 1996; Deković & Janssens, 1992; antisocial behaviour: Newcomb, Bukowski & Pattee, 1993; Veenstra et al., 2005). The chance of an error occurring due to a single reporter's experience with the child was significantly reduced. The frequency with which each participant was nominated was divided by the number of classmates who were nominating, in order to adjust for class size. Cronbach's alpha for antisocial and prosocial behaviour was .93 and .85, .92 and .84, .86 and .85, and .93 and .87 for the first, second, third and fourth occasion, respectively.

Participant roles in the bullying process. A sociometric questionnaire informed about the perceived participant roles in the bullying process of peers in the classroom, similar to nomination measures used in past research (e.g. Goossens, Olthof & Dekker, 2006). First, children were presented with an explanation of the concept of bullying and two questions to verify their understanding. If their answers were wrong, they received an explanation by the test administrator until they understood the concept of bullying. Then, the children were asked whether they were bullied. Depending on the answer to this question, the children were asked to name the classmate(s) that bullied them: "Who starts bullying you?", or to name the classmate(s) that bullied someone else in their classroom: "Who starts bullying?". A list of all their classmates was displayed in random order and they could click on the name(s) of their classmates that fit the description. The children could name an unlimited number of classmates. Pupils from other classes could also be mentioned using an open question. Further questions concerned joining the bully: "Who joins bullying you?" or: "Who joins bullying?", helping the victim: "Who helps you when you are bullied?" or: "Who helps the victim of the bullying?", and being victimized: "Who gets bullied?" in case the child was not bullied. Three different forms of bullying were distinguished and explained by drawings preceding the questions: (a) physical bullying (i.e., physical peer aggression, such as hitting, kicking or pushing); (b) verbal bullying (i.e., behaviours such as calling names or saying mean or unkind things and (c) relational bullying, a concept that referred mainly to social exclusion (e.g. Perren & Alsaker, 2006). Object-related bullying (e.g., taking away or breaking other child's belongings) and other specific forms of indirect bullying were left out, since it yielded least responses in a pilot study (conducted by Meijer, 2012). Instead, children could specify other forms of bullying in an open question format. All nominations were counted and divided by the number of participating classmates (and multiplied by 100) so that the data were comparable across classes. This yielded percentage scores (range 0-100) that indicated by what percentage of classmates each child was named as a bully, assistant, victim, outsider or defender. As such, peer nominations were aggregated across multiple nominators, which enhanced the reliability and validity of the data. Moreover, as insiders, peers can identify characteristics and relationships of children that are considered relevant from the perspective of those who ultimately determine a child's social status and integration in the peer group (Rubin, Bukowski & Parker, 1998).

Covariates age, gender, socioeconomic status, and scholastic ability. Age, gender, socioeconomic status, and scholastic ability were obtained from the school administration system and included as covariates. Socioeconomic status of the children was measured by averaging the level of highest completed education of their father and mother on a scale from 1 to 7 (1=kindergarten, 7=university+). Scholastic ability was measured by taking the average of nationally normed achievement tests on math and reading at the time of the first and second occasion. Scores range from 1 (lowest score) to 5 (highest score).

2.4 Analytical method

The effects of the intervention program were evaluated with the help of multilevel growth curve modelling with MLwiN (Rashbach, Charlton, Browne, Healy & Cameron, 2009) in order to account for the nested data structure. Three level models were fitted, with the first level representing change over time (within individual children), the second level representing individual differences between the children (within classes), and the third level representing differences between classes. Multilevel analysis takes into account this data structure by using a hierarchical linear model that allows for within group variability as well as between group variability (Snijders & Bosker, 2011). The differences between the four conditions on the outcome variables were examined separately after controlling for gender, age, socioeconomic status and scholastic ability. The predictors of interest were time-, individual- and class-specific and were included in the model as level 1, level 2 and level 3 predictors. The outcomes of interest were agreeableness, pro- and aggressive behaviour, bullying, assisting, defending, victimization, and outsider behaviour. In order to look at different growth curves, the four occasions of data collection were caught in a time variable with T1 = 0, T2 = 7 (months), T3 = 11 (months) and T4 = 17 months. The intervention was coded with three dummy variables; intervention versus control (only intervention between T1 and T2 = 1, control = 0), delayed intervention vs control I1 (only intervention between T3 and T4 = 1, control = 0), and extended intervention vs control I1 (intervention between T1 and T4 = 1, other = 0). For ease of interpretation as well as estimation, the level 1 and level 2 predictors were centred around the grand mean prior to statistical analysis.

The starting point of the multilevel growth models was the so-called empty model without any predictors. The empty models provided preliminary

information about the variance of the dependent variable between occasions within individual children (i.e. level 1), between individual children within classes (i.e. level 2) and between classes (i.e. level 3). Four forward steps after the empty models were distinguished: (1) adding different time variables at level 1; linear (time), quadratic (time^2), and cubic (time^3) growth representing different growth curves, (2) adding fixed predictors at level 2 (i.e. control variables); fixed effects do not vary across classes and can be regarded as the average effect over the whole population of children, (3) adding the fixed explanatory variables of main interest at level 3 (i.e. the intervention dummies) to evaluate the unique role in the prediction of the dependent variable while controlling for the variables entered in the previous model step, (4) adding interaction effects between time and intervention dummies (i.e. intervention*time variables) to examine the intervention effects by comparing the growth curves for the separate conditions, and (5) adding second-order interaction effects between time, intervention dummies and gender and age (i.e. boy*intervention*time and age*intervention*time variables) investigating whether the intervention effects differed depending on gender or age of the child. Having tested for random slopes and included those that were significant ($p<0.05$), the significance of both the fixed and interaction effects was evaluated with the t-test, based on the ratio of parameter estimate to standard error. Insignificant random slopes and interaction effects were removed from the model and this resulted in the final model. Comparisons between the deviance statistic of the final models and the deviance statistic of models with less parameters indicate that the final models we present were significantly the best fitting models for the given data.

We started by modelling the effects of the different intervention conditions on agreeableness. Since agreeableness might be a success-promoting factor of interventions targeting bullying-related behaviour in middle childhood, we first wanted to identify whether education was able to influence children's behavioural tendencies towards agreeableness. After that, we considered the effects the intervention had on bullying-related behaviour and pro- and antisocial behaviour. Then, after establishing all final models, we examined the role of agreeableness in predicting the other outcomes of the intervention. Since the intervention was aimed at promoting agreeableness, we were curious to see whether agreeableness mediated the effects the intervention had on the other outcomes. In order to get a hunch for these mediation effects, agreeableness was added to the equations of the

final models predicting the other outcomes. Smaller effects in the model would indicate that the effect runs through agreeableness.

3. Results

3.1 Missing data

The data contains some missing values, mainly due to unit non-response. The percentages of missing data were not high, respectively 2.9%, 2.5%, 17.6% and 8.7% at occasion 1, 2, 3 and 4. By occasion 2 one child entered the school and three children left. Between occasion 2 and 3 and occasion 3 and 4 respectively three and four children left. Furthermore, at occasion 3 one teacher did not fill in the questionnaire about the personality characteristics of the children, leading to the largest reduction of the number of children participating in the current inquiry. Nevertheless, the children who did not miss any data did not differ from these children in terms of gender, scholastic ability, socioeconomic status, agreeableness and prosocial behaviour at occasion 2 and 4, antisocial behaviour at occasion 1, 2 and 4, and the participant roles at occasion 2 ($\chi^2(1)=.33$; $p=.57$; $t(243)=-1.32$; $p=.19$; $t(243)=1.12$; $p=.26$; $t(239)=.74$; $p=.46$; $t(229)=-1.06$; $p=.29$; $t(242)=1.76$; $p=.08$; $t(236)=1.21$; $p=.22$; $t(242)=-1.67$; $p=.10$; $t(239)=1.46$; $p=.15$; $t(236)=.19$; $p=.84$; $\chi^2(4)=.67$; $p=.96$). However, non-participating children at occasion 3 were younger and had a higher mean score on agreeableness and prosocial behaviour at occasion 1 than participating children (respectively $t(244)=7.61$; $p<.01$, $d=1.51$; $t(239)=-.77$; $p<.01$, $d=.64$; $t(244)=-3.44$, $p<.01$, $d=.61$) and were underrepresented in the assistant participant role and overrepresented in the defender participant role at occasion 1 and 4 ($\chi^2(4)=15.0$; $p<.01$; $\chi^2(4)=14.4$; $p<.01$). Since the statistical model for multilevel repeated measures data does not require the same number of measurement occasions per individual, all of the available data was incorporated into the analysis. This means the data is unbalanced across time.

3.2 Descriptive statistics

At the first occasion, the four conditions did not significantly differ from one another with regard to gender, age, scholastic ability, and socioeconomic background ($F(3)=.03$; $p=.99$; $F(3)= 2.36$; $p=.07$; $F(3)= 2.04$, $p=.11$; $F(3)=.53$, $p=.66$ respectively). In Table 1 the means and standard deviations of all outcome variables are shown separately for the different intervention conditions and the

Table 1

Mean and standard deviations of the outcome variables for the different conditions for the four occasions

	Agreeableness	Antisocial behaviour	Prosocial behaviour	Bullying	Assisting	Defending	Victimization	Outsider behaviour
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Control								
T1	3.40 (.95)	15.73 (19.38)	46.26 (15.12)	1.85 (3.37)	1.12 (1.73)	3.48 (3.05)	4.10 (1.92)	89.45 (5.53)
T2	3.70 (.97)	14.08 (19.09)	43.51 (17.12)	1.68 (2.91)	.83 (1.25)	5.38 (3.70)	1.79 (2.70)	90.32 (4.90)
T3	3.78 1.00)	9.73 (17.09)	47.26 (16.78)	1.04 (2.02)	.97 (2.81)	3.59 (3.14)	.78 (1.82)	93.62 (7.68)
T4	3.60 (.89)	9.85 (18.18)	55.96 (19.03)	1.39 (2.33)	1.54 (2.32)	6.60 (3.49)	1.50 (2.17)	88.97 (6.66)
Intervention								
T1	3.51 (1.34)	20.59 (23.01)	48.34 (20.88)	2.74 (4.86)	1.86 (2.59)	3.87 (2.92)	2.84 (2.24)	88.68 (7.08)
T2	3.43 (1.25)	16.52 (18.53)	36.45 (18.27)	3.85 (7.04)	2.96 (4.14)	5.60 (4.30)	2.95 (3.42)	84.64 (12.81)
T3	3.43 (1.06)	19.74 (23.29)	45.20 (21.13)	3.60 (7.04)	2.60 (3.82)	4.10 (3.75)	1.82 (2.66)	87.87 (12.23)
T4	3.77 (1.07)	18.82 (24.66)	47.40 (20.79)	4.16 (6.81)	3.76 (3.90)	5.28 (5.69)	2.84 (4.07)	83.95 (12.19)
Delayed intervention								
T1	3.54 (1.03)	19.43 (22.51)	45.75 (21.02)	4.31 (7.02)	4.33 (4.99)	5.99 (4.74)	5.94 (3.59)	79.42 (12.28)
T2	3.54 (.96)	16.74 (21.07)	38.35 (18.92)	3.64 (5.56)	2.97 (4.30)	4.21 (3.07)	1.73 (2.28)	87.46 (10.60)
T3	3.44 (.93)	16.78 (22.07)	46.43 (14.19)	4.00 (4.81)	3.17 (2.56)	6.89 (5.07)	2.26 (2.98)	83.68 (10.94)
T4	3.76 (.96)	18.62 (22.30)	50.56 (13.96)	3.23 (3.72)	2.17 (2.55)	6.21 (3.27)	1.48 (2.73)	86.90 (7.74)
Extended intervention								
T1	2.77 (1.18)	19.33 (21.54)	55.75 (16.77)	4.82 (6.30)	5.23 (5.66)	4.21 (3.37)	4.39 (3.50)	81.35 (12.39)
T2	3.36 (1.20)	13.16 (19.99)	55.85 (18.58)	4.24 (6.35)	3.07 (3.77)	7.78 (4.30)	2.78 (4.65)	82.13 (11.20)
T3	3.58 (1.25)	14.44 (19.61)	66.39 (14.97)	3.58 (6.01)	3.94 (5.79)	10.08 (5.27)	2.74 (4.87)	79.67 (12.31)
T4	3.89 (1.08)	13.27 (20.17)	69.97 (16.51)	2.69 (4.48)	2.32 (3.14)	6.16 (3.67)	1.51 (2.04)	87.33 (6.82)

control condition at the four occasions. At the first measurement occasion, the four conditions did not significantly differ with regard to antisocial behaviour and bullying (respectively F(3)=.42; p=.81; F(3)= 2.50; p=.06). However, they did differ with regard to agreeableness, prosocial behaviour, assisting, defending, victimization and outsider behaviour (respectively F(3)=1707.94; p<.01; F(3)=1313.67; p<.01; F(3)=10.02; p<.01; F(3)=9.45; p<.01; F(3)= 20.65; p<.01; F(3)= 18.09, p<.01). Children in the extended intervention scored lower on agreeableness and outsider behaviour, and higher on prosocial behaviour and assisting in comparison to the other conditions. Children in the delayed intervention condition scored higher on assisting, victimization and defending and lower on outsider behaviour. The intervention group scored lower on victimization than the other conditions.

Comparing the intervention groups means over time, several positive trends can be noted. The biggest change took place in the mean of agreeableness, for which a substantial increase occurred in the extended intervention (from 2.77 to 3.89), with a much smaller change in the control condition (from 3.40 to 3.60). Likewise, there was a change favouring the extended intervention in all the outcomes from occasion 1 to occasion 4, albeit some differences were small (e.g., for defending). For the children receiving the intervention, positive and negative

Table 2

Variance estimates and intraclass correlations (ICC) of the outcome variables

	Variances					
	σ^2_e	σ^2u	σ^2v	ICC1	ICC2	ICC3
Agreeableness	.614 (.033)	.598 (.070)	.000 (.000)	.51	.49	.00
Antisocial behaviour	88.861 (4.673)	364.642 (35.258)	.000 (.000)	.20	.80	.00
Prosocial behaviour	146.248 (7.690)	182.543 (20.371)	62.442 (32.197)	.37	.47	.16
Bullying	9.395 (.493)	20.504 (2.112)	.657 (.716)	.31	.67	.02
Assisting	7.037 (.370)	6.868 (.803)	1.106 (.657)	.47	.46	.07
Defending	13.473 (.707)	1.698 (.503)	3.006 (1.423)	.74	.09	.17
Victimization	8.474 (.445)	1.937 (.393)	.636 (.360)	.77	.18	.06
Outsider behaviour	67.509 (3.021)	40.114 (5.102)	16.244 (8.243)	.55	.32	.13

Note. σ^2_e= variance between occasions within children; σ^2u variance between children within classes; σ^2v = variance between classes. ICC1, ICC2, and ICC3 = proportion of variance at the time, individual, and class level.

changes were found in between occasion 1 and 2. For the delayed intervention, mainly positive changes were found in between occasion 3 and 4.

3.3 Multilevel models

The multilevel empty model was estimated to obtain the intra-class correlation coefficients (ICC) indicating the proportion of variance accounted for at the individual and class level. For each dependent variable, the variance was estimated at three levels: time, individuals, and classes (see Table 2). Overall the time- and child-level variance was higher than the class-level variance. This indicates that individuals and time are more important for variation in the outcome measures than classes. The highest proportion of variance associated with time were obtained for defending (ICC1=.74) and victimization (ICC1=.77). For individual factors the highest proportion of variance were obtained for antisocial behaviour (ICC2=.80) and bullying (ICC2=.67). Between class variance was highest for defending and prosocial behaviour (respectively ICC3= .17 and ICC3=.16).

3.3.1 Intervention effects on agreeableness

The intervention effects on agreeableness are reported in Table 3 and visualized in Figure 1. The linear time effect of the intervention condition, Intervention*Time, was b= -.083; p <0.01. This indicates that agreeableness decreases over time when comparing the intervention with the control condition. However, in combination with the significant, but small quadratic time effect, Intervention* Time^2, of b=.006; p <0.01 the linear decrease is transformed into a slight parabolic

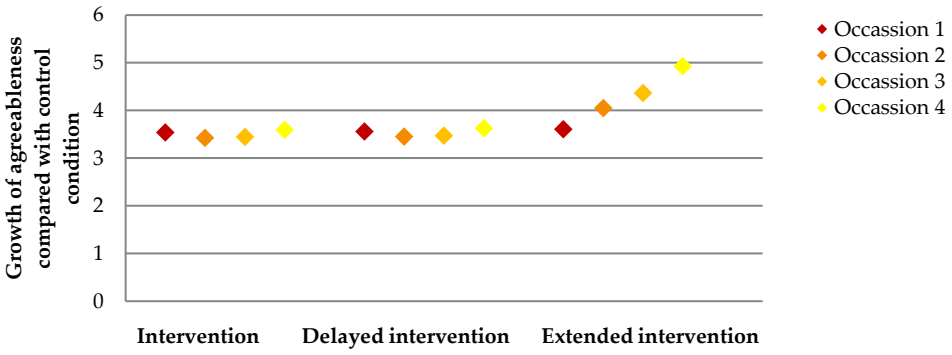


Figure 1. Growth of agreeableness in comparison with the control condition for the different intervention conditions over time

shape that opens upward. The cubic time effect of the intervention condition was not significant and was left out of the model. When combining the linear and quadratic effect of the intervention, children in the intervention condition decreased in agreeableness during the first year of the intervention and slightly increased in agreeableness during the second year, compared to the control condition.

For the delayed intervention condition a similar pattern was found. The linear time effect was significant and negative (b=-.107; p <0.01), the quadratic time effect was significant and positive (b=.006; p <0.01), and the cubic effect was not significant. This indicates that children in the delayed intervention decreased in agreeableness during the first year of the intervention and a slightly increased in agreeableness during the second year, compared to the control condition.

Compared to the control condition, the extended intervention classes show a significant quadratic time effect of b=.005; p <0.01. The linear and cubic time effects were not significant. This means that the growth of the extended intervention classes compared to the control condition has a parabolic shape that opens upward. Thus, agreeableness decreased during the first year and increased during the second year of the extended intervention.

3.3.2 Differential intervention effects on agreeableness

Differential intervention effects on agreeableness depending on gender were only found when comparing the control to the intervention classes. Differences between girls and boys were found with regard to the linear time effect of the intervention condition; the three-way effect of Intervention* Time*Girl was b=-.049; p <0.01. As Figure 2 shows, this means that for boys in the intervention classes slightly increased in agreeableness over time, whereas girls decreased in agreeableness over time in comparison to the control group.

Three-way interaction effects with age were found for the intervention and extended intervention condition compared to the control condition. First, we found three-way interaction effects indicating that the growth in agreeableness of children in the intervention classes depend on age (Intervention* Time*Age was b=.071; p <0.01 and Intervention* Time^2*Age was b=-.004; p <0.01). A visualization of this interaction effect can be found in Appendix F, Figure 1. The combination of the positive linear and negative quadratic effect indicates that the growth effect – a parabola that opens upward – changes according to age. When children were younger this growth effect mainly applied to the first year of the intervention. Thus,

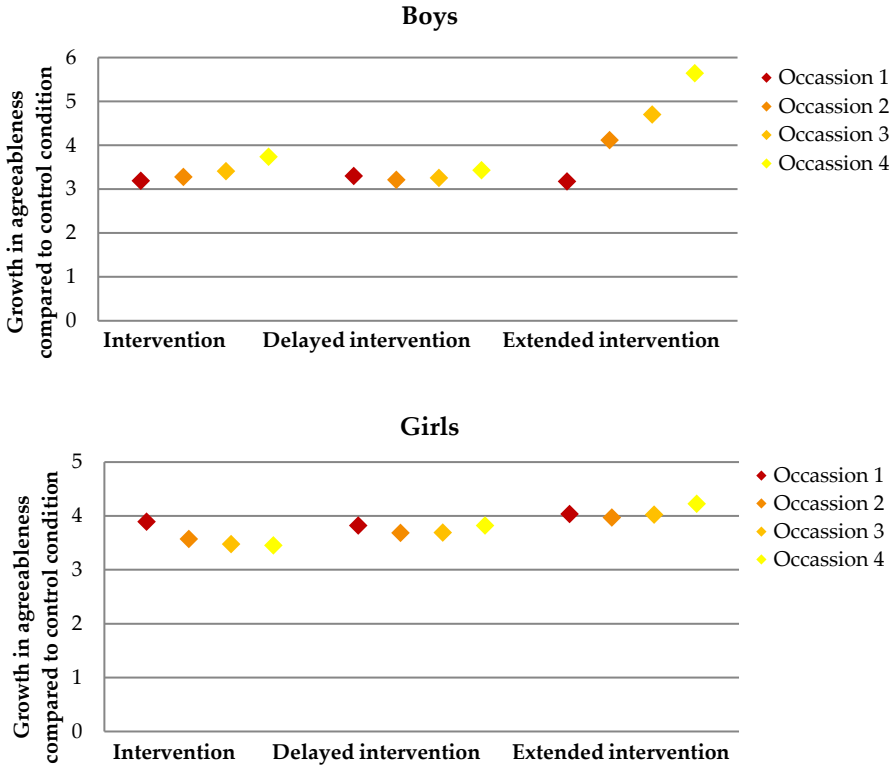


Figure 2. Growth in agreeableness in comparison with the control condition for the different intervention conditions over time for boys and girls separately

for younger children agreeableness decreases in the first year and slightly in the second year of the intervention, in comparison to the control condition. When children were older this growth effect becomes opposite. Thus, agreeableness increases during the first year of the intervention and slightly in the second year in comparison to the control condition. Second, two three-way interactions with age were found for the growth of the extended intervention condition compared to the control condition. A visualization of this interaction effect can be found in Appendix F, Figure 2. Overall the extended intervention increased agreeableness in comparison to the control condition. When children were younger, the extended intervention decreased agreeableness during the first and second year. When children were older, the opposite effect was found (linear and quadratic three- way effects were b=.151; p <0.01 and b=-.007; p <0.01 respectively).

3.3.3 Intervention effects on the other outcomes

The intervention effects on the other outcomes are reported in Table 3 and Table 4. Particularly, the effects on antisocial behaviour, prosocial behaviour, bullying, assisting, defending, victimization and outsider behaviour are reported. We will now discuss the intervention effects on the aforementioned behavioural and bullying-related outcomes, starting with the effects of the intervention. After that we will continue discussing the effects of the delayed intervention followed by the effects of the extended intervention. First of all, children in the intervention classes, compared to the control classes, showed a linear decrease in antisocial and prosocial behaviour (b=-3.516; p <0.01 and b=-4.373; p <0.01 respectively) that slows down during the second year because of the significant quadratic time effect (b=.574; p <0.01 and b=.652; p <0.01 respectively) in combination with the cubic time effect (b=-.021; p <0.01 and b-.026; p <0.01 respectively). A negative cubic growth effect first shows a decline, than levels off, again followed by a decline.

When comparing the delayed intervention to the control condition, defending slightly decreased during the first and slightly increased during the second year (linear, quadratic and cubic effects were b=-2.933; p <0.01, b=.472; p <0.01 and b=-.018; p <0.01 respectively). Victimization decreased over time when comparing the delayed intervention to the control condition (linear, quadratic and cubic effects were b=-1.214; p <0.01, b=.172; p <0.01 and b=-.006; p <0.01 respectively). Outsider behaviour increased over time when comparing the delayed intervention to the control condition, especially during the first year (linear, quadratic and cubic effects were b=5.592; p <0.01, b=-.836; p <0.01 and b=.032; p <0.01 respectively).

Compared with the control condition, the children in the extended intervention classes showed a decrease in antisocial and prosocial behaviour, and assisting with a stabilization in the middle of the occasions (linear, quadratic and cubic effects were b=-3.458; p <0.01, b=.545; p <0.01, and b=-.020 p <0.01; b=-3.337; p <0.01, b=.630; p <0.01, and b=-.019; p <0.01; and b=-2.668; p <0.01, b=.406; p <0.01 and b=-.016; p <0.01 respectively). For defending there was a slight decrease over time for the extended intervention versus control condition (linear, quadratic and cubic effects were b=-1.750; p <0.01, b=.432; p <0.01, b=-.020; p <0.01), whereas outsider behaviour increased over time (linear, quadratic and cubic effects were b=4.761; p <0.01, b=-.893; p <0.01, b=.039; p <0.01).

3.3.4 Differential intervention effects on the other outcomes depending on gender

Differences between girls and boys were found with regard to the effects of the different intervention conditions. First of all, children in the intervention classes, compared to the control classes, showed a decline in antisocial behaviour over time that levels off in between in the first and second year of the intervention, but continues during the second year. Interestingly, this decline in antisocial behaviour during the first and second year was apparent for boys, but not for girls (linear three-way effect was b=.642; p <0.01, for a visualization see Figure 3, Appendix F).

Second, the delayed intervention condition increased outsider behaviour over time compared to the control condition, especially during the first year. However, boys' outsider behaviour increased more over time than girls' outsider behaviour (linear three-way effect was b=-.427; p <0.01, see Figure 4, Appendix F).

Third, for the extended intervention condition several three-way effects with gender were found. Children in the extended intervention did not show significant differences in bullying over time in comparison to the control group. However, boys showed a slight decrease in bullying when comparing the extended intervention with the control condition whereas girls showed a slight increase (linear three-way effect was b=.195; p <0.01, see Figure 5, Appendix F). Additionally, compared with the control condition, the children in the extended intervention classes showed a decrease in assisting with a stabilization in the middle of the occasions. However, compared with the control group, boys in the extended intervention classes showed bigger decreases in assisting than girls (linear, quadratic and cubic three-way effects were b=2.044; p <0.01, b=-.309; p <0.01, b=.012; p <0.01, see Figure 6, Appendix F). In line with this, outsider behaviour generally increased over time when comparing the extended to the control condition. This effect was visible for boys and girls, but for boys this effect was only slightly stronger (linear three-way effect was b=-.455; p <0.01, see Figure 7, Appendix F). Furthermore, overall victimization did not change over time for the extended intervention versus control condition. However, among girls victimization showed a strong decrease, whereas among boys victimization remained stable (linear, quadratic and cubic three-way effects were b=-2.094; p <0.01, b=.326; p <0.01, b=-.012; p <0.01, see Figure 8, Appendix F).

3.3.5 Differential intervention effects on the other outcomes depending on age

Three differential effects were found depending on age for the intervention versus control. First, prosocial behaviour shows a decline in the first year and that levels off in the following year when comparing the intervention with the control condition. This pattern is slightly strengthened when children were younger and slightly weakened when children were older (linear, quadratic and cubic three-way effects were b=-2.460; p <0.01, b=-.432; p <0.01, b=-.016; p <0.01, for a visualization see Figure 9, Appendix F). Second, the intervention condition showed no effect on defending and outsider behaviour over time in comparison to the control condition. However, defending appeared to increase over time when children were older and decrease over time when children were younger (linear, quadratic and cubic three-way effects were b=1.141; p <0.01, b=.165; p <0.01, see Figure 10, Appendix F). Outsider behaviour showed a strong decrease in the first year for younger children, with a continuing decrease in the second year. For older children a similar, but slightly less strong pattern for outsider behaviour was found when comparing the intervention to the control condition (b=-.006; p <0.01; and linear three-way effect was b=1.354; p <0.01, see Figure 11, Appendix F).

The effects of the delayed intervention also varied with age. First, the delayed intervention did not show an effect on bullying when compared with the control condition. When children were older, however bullying decreased during the first year and the second year. When children were younger the opposite effect was found (linear and quadratic three-way effects were b=-.717; p <0.01. b=.032; p <0.01, see Figure 12, Appendix F). Second, the delayed intervention did not show a significant decrease of assisting when compared with the control condition. Yet, assisting showed a slightly stronger decreased over time during the first year when children were younger and a slightly stronger decreased over time during the second year when children were older (cubic three-way effect was b=.004; p <0.01, see Figure 13, Appendix F). Last, defending decreased during the first year and leveling off during the second year when comparing the delayed intervention to the control condition. When children were younger this effect was slightly strengthened (linear, three-way effect was b=-1.066; p <0.01, see Figure 14, Appendix F).

Differential effects depending on age were also visible in the extended intervention. Specifically, these effects were present for assisting, victimization, and outsider behaviour. Compared with the control condition, the children in the extended intervention classes showed a decrease in assisting with a stabilization in

the middle of the occasions. These decreases were strengthened when children were younger (linear, quadratic and cubic three-way effects were *b*=2.131; *p* <0.01. *b*=-.357; *p* <0.01 and *b*=.014; *p* <0.01 respectively, see Figure 15, Appendix F). Generally, the extended intervention did not change victimization over time when compared to the control condition. However, when children were younger victimization decreased over time. When children were older victimization increased over time (linear, quadratic and cubic three-way effects were *b*=1.709; *p* <0.01, *b*=-.232; *p* <0.01, *b*=.008; *p* <0.01, see Figure 16, Appendix F). Last, outsider behaviour increased over time when comparing the extended and control condition. This increase was strengthened within the first and second year when children were younger. When children were older, outsider behaviour slightly decreased within the first and second year (linear, quadratic and cubic three-way effects were *b*=-4.211; *p* <0.01, *b*=.754; *p* <0.01, *b*=-.031; *p* <0.01, see Figure 17, Appendix F).

Table 3

Final Multilevel Models: Intervention Effects for Agreeableness, and Antisocial and Prosocial Behaviour

	Agreeableness	Antisocial behaviour	Prosocial behaviour
	Estimate (S.E.)	Estimate (S.E.)	Estimate (S.E.)
Fixed effects			
Intercept	3.452 (.155)	20.032 (3.466)	41.442 (3.223)
Level 1 (time) variables			
Time	**.081 (.027)**	-.540 (1.203)	-2.025 (1.161)
Time*Time	**-.004 (.001)**	-.049 (.194)	.264 (.187)
Time*Time*Time		.003 (.007)	-.006 (.007)
Level 2 (individual) variables			
Girl	.310 (.215)	**-13.062 (4.797)**	**11.993 (1.490)**
Age	-.104 (.053)	-1.078 (0.753)	-1.693 (1.490)
Socioeconomic status	.072 (.101)	-2.251 (1.323)	**2.213 (.906)**
Scholastic ability	**.186 (.049)**	**-4.985 (1.092)**	**5.263 (1.144)**
Level 3 (class) variables			
Intervention	-.264 (.195)	4.431 (4.369)	-1.478 (4.058)
Delayed intervention	-.157 (.197)	**8.940 (4.400)**	-1.945 (4.014)
Extended intervention	-.280 (.229)	2.350 (4.968)	**16.129 (4.594)**
Cross-level interactions			
Intervention*Time	**-.083 (.036)**	**-3.516 (1.409)**	**-4.373 (1.502)**

Delayed intervention*Time	**-.107 (.035)**	-1.927 (1.416)	-2.359 (1.510)
Extended intervention*Time	.047 (.042)	**-3.458 (1.581)**	**-3.337 (1.801)**
Intervention*Time^2	**.006 (.002)**	**.574 (.227)**	**.652 (.242)**
Delayed intervention*Time^2	**.006 (.002)**	.350 (.228)	.371 (.244)
Extended intervention*Time^2	**.005 (.002)**	**.545 (.254)**	**.630 (.291)**
Intervention*Time^3		**-.021 (.009)**	**-.026 (.009)**
Delayed intervention*Time^3		-.013 (.009)	-.015 (.009)
Extended intervention*Time^3		**-.020 (.010)**	**-.025 (.011)**
Time*Girl	-.009 (.016)	**3.241 (1.005)**	
Time^2*Girl		**-.432 (.160)**	
Time^3*Girl		**.015 (.006)**	
Time*Age	-.010 (.013)		**3.630 (.585)**
Time^2*Age	.001 (.001)		**-.551 (.094)**
Time^3*Age			**.019 (.004)**
Intervention*Girl	.393 (.277)	.459 (6.223)	
Delayed intervention*Girl	.216 (.279)	-8.406 (6.203)	
Extended intervention*Girl	.429 (.312)	-4.375 (6.931)	
Intervention*Age	**.037 (.072)**		1.732 (2.024)
Delayed intervention*Age	.138 (.108)		2.613 (2.656)
Extended intervention*Age	-.108 (.183)		-.212 (3.720)
Cross-level three-way interactions			
Intervention*Time*Girl	**-.049 (.021)**	**.642 (.266)**	
Delayed intervention*Time*Girl	.001 (.002)	-.134 (.267)	
Extended intervention*Time*Girl	.012 (.023)	-.033 (.297)	
Intervention*Time^2*Girl			
Delayed intervention*Time^2*Girl			
Extended intervention*Time^2*Girl			
Intervention*Time^3*Girl			
Delayed intervention*Time^3*Girl			
Extended intervention*Time^3*Girl			
Intervention*Time*Age	**.071 (.018)**		**-2.460 (.786)**
Delayed intervention*Time*Age	.039 (.026)		1.314 (1.177)
Extended intervention*Time*Age	**.151 (.045)**		1.565 (2.019)
Intervention*Time^2*Age	**-.004 (.001)**		**.432 (.127)**
Delayed intervention*Time^2*Age	.001 (.001)		-.128 (.190)
Extended intervention*Time^2*Age	**-.007 (.003)**		-.019 (.326)
Intervention*Time^3*Age			**-.016 (.005)**
Delayed intervention*Time^3*Age			.004 (.007)

Extended intervention*Time^3*Age			-.004 (.013)
Variance components			
Classroom level variance	.000 (.000)	.000 (.000)	14.642 (9.132)
- Slope for Scholastic ability			7.185 (5.650)
- Slope for SES	.065 (.044)		
- Covariance	.000 (.000)		-2.842 (5.133)
Individual level variance	.417 (.051)	263.365 (25.668)	107.587 (12.392)
Time level variance	.478 (.026)	79.241 (4.178)	90.493 (4.769)
Deviance	2302.644	7604.609	7532.116

Note. In bold: significant differences using p<0.05.

Table 4

Final Multilevel Models: Intervention Effects for Bullying, Assisting, Defending, Victimization and Outsider Behaviour

	Bullying	Assisting	Defending	Victimization	Outsider Behaviour
	Estimate (S.E.)	Estimate (S.E.)	Estimate (S.E.)	Estimate (S.E.)	Estimate (S.E.)
Fixed effects					
Intercept	2.588 (.874)	2.047 (.570)	3.933 (.329)	2.025 (.406)	89.368 (1.400)
Level 1 (time) variables					
Time	-.006 (.110)	-.060 (.397)	**1.301 (.379)**	-.183 (.422)	-1.292 (.770)
Time*Time	.003 (.006)	-.001 (.064)	**-.214 (.061)**	-.014 (.068)	**.285 (.124)**
Time*Time*Time		.000 (.002)	**.009 (.002)**	.001 (.003)	**-.013 (.005)**
Level 2 (individual)					
Girl	-1.823 (1.208)	**-1.537**	**1.707 (.271)**	.173 (.561)	1.553 (1.937)
Age	-.238 (.305)	-.323 (.222)	**-.528 (.150)**	-.241 (.142)	**1.334 (.489)**
Socioeconomic status	-.106 (.335)	-.219 (.187)	.081 (.164)	-.155 (.156)	.430 (.538)
Scholastic ability	**-.957 (.276)**	**-.582 (.154)**	.207 (.137)	-.206 (.129)	**1.542 (.443)**
Level 3 (class) variables					
Intervention	**2.967 (1.100)**	**2.273 (.724)**	-.185 (.385)	.127 (.511)	**-4.851 (1.764)**
Delayed intervention	**2.948 (1.109)**	**2.489 (.726)**	**.976 (.386)**	**1.166 (.515)**	**-8.020 (1.779)**
Extended intervention	**3.790 (1.303)**	**4.401 (.852)**	**2.956 (.457)**	.891 (.605)	**-11.648 (2.088)**
Cross-level two-way					
Intervention*Time	.176 (.141)	.936 (.500)	-.145 (.490)	.724 (.531)	-1.314 (.996)
Delayed*Time	-.067 (.142)	-.425 (.508)	**-2.933**	**-1.214**	**5.592 (1.002)**

Extended*Time	-.022 (.167)	**-2.668**	**-1.750**	.224 (.629)	**4.761 (1.182)**
Intervention*Time^2	-.004 (.008)	-.134 (.081)	.037 (.079)	-.081 (.086)	.125 (.160)
Delayed*Time^2	-.001 (.008)	.053 (.082)	**.472 (.080)**	**.172 (.087)**	**-.836 (.161)**
Extended*Time^2	-.012 (.009)	**.406 (.096)**	**.432 (.094)**	-.006 (.102)	**-.893 (.191)**
Intervention*Time^3		.005 (.003)	-.002 (.003)	.002 (.003)	-.003 (.006)
Delayed*Time^3		-.002 (.003)	**-.018 (.003)**	**-.006 (.003)**	**.032 (.006)**
Extended*Time^3		**-.016 (.004)**	**-.020 (.004)**	-.000 (.004)	**.039 (.007)**
Time*Girl	-.046 (.065)	-.045 (.550)		-.053 (.584)	.125 (.145)
Time^2*Girl		.014 (.089)		-.030 (.094)	
Time^3*Girl		-.001 (.003)		.002 (.004)	
Time*Age	-.012 (.052)	-.093 (.139)	**1.157 (.191)**	-.204 (.147)	**-.810 (.386)**
Time^2*Age	.001 (.003)	.013 (.022)	**-.166 (.031)**	.029 (.024)	.117 (.062)
Time^3*Age		-.001 (.001)	**.006 (.001)**	-.001 (.001)	**-.004 (.002)**
Intervention*Girl	-2.658 (1.568)	**-1.963**		.645 (.729)	3.261 (2.514)
Delayed *Girl	-1.482 (1.569)	-1.235		-.900 (.729)	4.562 (2.516)
Extended*Girl	**-3.555 (1.747)**	**-3.696**		.284 (.813)	**6.218 (2.804)**
Intervention*Age	-.231 (.408)	.106 (.299)	**-.678 (.201)**	**-.438 (.190)**	**1.304 (.655)**
Delayed*Age	-.257 (.617)	-.096 (.426)	**-.730 (.302)**	-.421 (.287)	**1.449 (.990)**
Extended*Age	-.376 (1.025)	-.847 (.656)	**-1.422**	-.876 (.479)	**3.594 (2.516)**
Cross-level three-way					
Intervention*Time*Girl	.010 (.084)	-.743 (.714)		-.099 (.758)	-.173 (.188)
Delayed*Time*Girl	.111 (.085)	-.139 (.718)		-.104 (.763)	**-.427 (.189)**
Extended*Time*Girl	**.195 (.094)**	**2.044 (.794)**		**-2.094**	**-.455 (.209)**
Intervention*Time^2*Girl		.120 (.115)		.031 (.122)	
Delayed*Time^2*Girl		.016 (.116)		.037 (.123)	
Extended*Time^2*Girl		**-.309 (.128)**		**.326 (.136)**	
Intervention*Time^3*Girl		-.005 (.004)		-.001 (.005)	
Delayed*Time^3*Girl		-.000 (.004)		-.002 (.005)	
Extended*Time^3*Girl		**.012 (.005)**		**-.012 (.005)**	
Intervention*Time*Age	-.110 (.071)	-.057 (.187)	**1.141 (.257)**	-.114 (.199)	**1.354 (.519)**
Delayed*Time*Age	**-.717 (.106)**	.103 (.281)	**-1.066**	.572 (.299)	-.629 (.777)
Extended*Time*Age	-.193 (.177)	**2.131 (.471)**	-.140 (.646)	**1.709 (.501)**	**-4.211 (1.309)**
Intervention*Time^2*Age	.004 (.004)	.002 (.030)	**.165 (.041)**	.027 (.032)	-.185 (.084)
Delayed*Time^2*Age	**.032 (.006)**	-.084 (.046)	.089 (.062)	-.082 (.048)	.122 (.126)
Extended*Time^2*Age	.015 (.010)	**-.357 (.077)**	-.065 (.105)	**-.232 (.081)**	**.754 (.213)**
Intervention*Time^3*Age		-.000 (.001)	**-.006 (.002)**	-.001 (.001)	.006 (.003)
Delayed*Time^3*Age		**.004 (.002)**	.002 (.002)	.003 (.002)	-.008 (.005)
Extended*Time^3*Age		**.014 (.003)**	.004 (.004)	**.008 (.003)**	**-.031 (.008)**

Variance components					
Classroom level variance	.000 (.000)	.177 (.183)	.000 (.000)	.000 (.000)	.000 (.000)
Individual level variance	15.964 (1.629)	4.240 (.515)	1.940 (.417)	2.433 (.360)	36.240 (4.222)
Time level variance	7.933 (.418)	5.089 (.268)	9.650 (.508)	5.743 (.302)	39.486 (2.090)
Deviance	5282.923	4679.601	5079.357	4674.940	6673.371

Note. In bold: significant differences using p<0.05.

3.3.6 Mediating effect of agreeableness

The role of agreeableness in predicting the outcomes of the intervention was present, but relatively small. When agreeableness was added to the equations of the final models, no changes occurred in the intervention-related effects on bullying and defending. However, for antisocial behaviour several intervention-related effects became nonsignificant (the Extended intervention*Time effect became *b*=-2.761, *S.E.*=1.506, Extended intervention*Time^2 became *b*=.416, *S.E.*=.243 and Extended intervention*Time^2 became *b*=-.014, *S.E.*=.009). As indicated before, smaller effects would indicate that the intervention effects on other outcomes, in this case antisocial behaviour, runs through agreeableness. Similarly, five intervention effects (Extended intervention*Time became *b*=2.824, *S.E.*=1.794, Extended intervention*Time^3 became *b*=-.021, *S.E.*=.011, Intervention*Time*Age became *b*=-1.386, *S.E.*=.137, Intervention*Time^2*Age became *b*=.253, *S.E.*=.818, and Intervention*Time^3*Age became *b*=-.008, *S.E.*=.005) turned insignificant in the model predicting prosocial behaviour when adding agreeableness. This means that part of effects of the intervention and extended intervention on prosocial behaviour as compared to the control condition might be due to a change in agreeableness. Additionally, in the model predicting victimization three effects were not present anymore when adding agreeableness (Delayed intervention*Time^2* became *b*=.152, *S.E.*=.088 and Extended intervention*Time*Girl became *b*=-1.613, *S.E.*=.863 and Intervention*Time^2*Girl became *b*=.261, *S.E.*=.140). For the model predicting outsider behaviour, the Intervention*Time*Age effect became insignificant (*b*=.515, *S.E.*=.542) and for the model predicting assisting the Delayed intervention*Time^3*Age became insignificant (*b*=.003, *S.E.*=.002). As expected, no insignificant intervention effects turned significant when adding agreeableness to the equation.

4. Discussion

The aim of this paper was to examine how education, one of the most important contextual factors in childhood, might be able to influence children's behavioural tendencies towards agreeableness in order to target bullying-related behaviour. Based on previous empirical research, we argued that agreeableness might be a success-promoting factor of interventions targeting bullying-related behaviour in middle childhood. By solely focussing on agreeableness as a potential underlying success-promoting factor of interventions aimed at affecting anti- and prosocial behaviour, we tried to get more information about the processes explaining the effectiveness of interventions (Does stimulating agreeableness lead to a reduction of antisocial behaviour and an increase in prosocial behaviour?) instead of solely focussing on evidence of effectiveness (Does the intervention reduce antisocial behaviour and increase prosocial behaviour?) (e.g. Gravemeijer & Kirschner, 2007). Thus, we were curious to know whether we would be able to influence agreeableness and moreover, whether an increase in agreeableness might benefit other behavioural outcomes in middle childhood. Specifically, this paper investigates the effects of a class-based intervention program promoting agreeableness on agreeableness as well as antisocial and prosocial behaviour and bullying-related behaviour. These effects were examined by comparing children in a control condition with three intervention conditions at the beginning and at the end of two school years: (a) the intervention condition in which the intervention was implemented during the first schoolyear, (b) the delayed intervention condition in which the intervention was implemented during the second schoolyear, and (c) the extended intervention condition in which the intervention was implemented during both school years. As expected, promising effects of the class-based intervention were found for all conditions fostering agreeableness and prosocial behaviour and reducing bullying-related behaviour and antisocial behaviour.

An encouraging and important result is that we succeeded in the stimulation of agreeableness in all intervention conditions. Moreover, the time frame in which agreeableness increased varied per condition. Within the delayed intervention condition agreeableness increased during the second year, when the intervention was implemented. Within the intervention condition agreeableness also increased during the second schoolyear, but also during the first schoolyear for older boys. Within the extended intervention condition agreeableness increased

during both years for older children. These results support the idea that agreeableness is a malleable construct and is susceptible to change in light of environmental inputs in education (e.g. Bergeman et al., 1993; Graziano & Eisenberg, 1997). Moreover, this demonstrates that our approach for manipulating agreeableness in primary education has potential. Our focus on the promotion of self-regulation of (negative) emotions and on the promotion of positive interpersonal relations targeting the lower order traits of agreeableness appeared to be an adequate method for promoting agreeableness in children (e.g. Eisenberg, Spinrad & Eggum, 2010; Jones & Jones, 2010; Warden & Mackinnon, 2003).

Turning to the intervention effects on the other behavioural outcomes, we found that the positive effects of the intervention were especially visible in the extended intervention condition. The extended intervention not only induce an increase in agreeableness, it also prompted a decline in antisocial behaviour, bullying (for older boys), assisting (especially for younger boys), and victimization (for younger girls). Moreover, outsider behaviour increased, especially for younger boys. An unexpected outcome is the decrease of prosocial behaviour and defending over time. Additionally, the delayed intervention and the intervention condition disclosed some positive effects. In the delayed intervention condition assisting and victimization decreased and outsider behaviour increased compared to the control condition. Assisting mainly decreased among younger children during the first year and older children during the second year. Contrarily, the delayed intervention showed a decrease in defending. No other significant effects were found. Then, the intervention condition showed mixed effects, with a small majority of positive effects. The intervention induced a decrease in antisocial behaviour (for boys) and prosocial behaviour (slightly more for younger children) and an increase in defending (among older children) and outsider behaviour (for older children). No effects were found on bullying, assisting and victimization.

Thus, the effects found on antisocial behaviour, bullying, assisting, victimization and outsider behaviour all favoured the intervention conditions versus the control condition. This is in line with our hypotheses that increasing agreeableness might decrease antisocial behaviour and bullying-related behaviour (Cook et al., 2010; McCrae & John, 1991; Shiner, 2006). Unexpectedly, however, prosocial behaviour and defending decreased over time in all conditions. Indeed, research seems to agree that antisocial and prosocial behaviour are not the opposite

ends of a general construct on social behaviour; instead, they are two separate constructs (Krueger, Hicks & McGue, 2001). Therefore, a reduction in antisocial outcomes does not imply an increase in prosocial outcomes. However, defending did increase for older children in the intervention condition. An explanation for this increase in defending is that being prosocial in a specific context (i.e. defending) might be different from being prosocial in general (i.e. helping and being nice), especially for younger children (e.g. Grusec & Redler, 1980). Overall, however, we did not find support for the hypothesis that stimulating agreeableness increases defending and prosocial behaviour (e.g. Graziano & Eisenberg, 1997; Sneed, 2002; Tani et al., 2003).

In order to test our hypotheses that increasing agreeableness might decrease antisocial behaviour and bullying-related behaviour and increase prosocial behaviour, agreeableness was added to the equations of the final models predicting the other outcomes. In the models predicting antisocial behaviour, assisting, victimization, outsider behaviour, and prosocial behaviour several intervention-related effects turned insignificant. No changes occurred in the intervention-related effects on bullying and defending. This indicates that some effects of the intervention might run through agreeableness. However, the role of agreeableness in predicting the other outcomes of the intervention was relatively small. Therefore, we are not sure whether the changes in agreeableness were caused the changes in the other outcomes. However, there are some indications that stimulating agreeableness in primary education has the potential to redirect negative trajectories of functioning. A replication of the present study would provide valuable evidence for this claim. Additionally, studying success-promoting factors and underlying processes of interventions targeting bullying-related behaviour in middle childhood remains an important topic for future research.

Generally, the above findings suggest that the effectiveness of interventions grows when carried out over a longer period of time. Changing behavioural tendencies in a class environment might require patience and persistence. Unfortunately, however, most efforts to improve the social environments of schools are not sustained (e.g. Supovitz & Weinbaum, 2008). Since we partly controlled the implementation of the intervention program by letting an external teacher teach the lessons, it was relatively easy to continue the intervention for a longer time and to implement the program as intended. Moreover, the external teacher not only turned

out to be an advantage for the sustainability of the program, but also for the implementation of the program by the teachers later on. By modelling the lessons, the teachers could more easily teach the lessons themselves by imitating the external teacher. According to social learning theory, modelling is a powerful instrument for learning (e.g. Bandura, 1986). Even though we did not assess this, it might be the case that a change in teacher behaviour affected the behaviour of the children in our study instead of, or in combination with, the direct effect of the program directed at agreeableness. Teachers might have become more attentive to prosocial and antisocial behaviour of the children in their class. Future studies could look into this mediating effect of teacher behaviour on the effectiveness of an intervention program aimed at affecting pro- and antisocial behaviour. Furthermore, it would be interesting to examine this effect for an intervention program modelled by an external teacher versus an intervention program implemented in a less energy-consuming way, for example with teacher manuals and training days.

As previous studies have shown that gender and age are important predictors of bullying-related behaviour as well as anti- and prosocial behaviour (e.g. Salmivalli & Voeten, 2004; Whitney & Smith, 1993), we looked into the differential effects of age and gender on the effectiveness of the intervention conditions. A striking number of gender specific effects were found. Since antisocial behaviour, bullying, and assisting were more frequently reported for boys than for girls, this might explain that a decrease in these outcomes was more apparent among boys than girls in the intervention and extended intervention condition. The same applies to the decrease in victimization among girls. Moreover, the decrease in bullying and assisting for boys clarifies the increase in outsider behaviour among boys in the intervention and extended intervention. Research has shown that peers, especially girls, are more likely to nominate girls as being prosocial and nominate boys as being bullies (Warden, Cheyne, Christie, Fitzpatrick & Reid, 2003; Warden & Mackinnon, 2003). Fewer differences have been found in observational studies (Fabes, Martin & Hanish, 2002; Zahn-Waxler et al. 2001). This has led researchers to believe that gender differences in other reported behaviour may reflect people's conceptions of what boys and girls are supposed to be like rather than how they actually behave. A future study examining the effects of a class-based intervention promoting agreeableness and other behavioural outcomes might thus benefit from observational data.

As expected, age also played a major role in the effectiveness of the three different intervention conditions. However, the results were far from consistent. This led us to the idea that another characteristic of the classroom might be a confounding variable. Since the classes are grade-specific, differences in age directly reflect differences between classes and therefore differences in teachers. As teachers are an important player in children's social behaviours (Davis, 2003; Jennings & Greenberg, 2009), teachers might have caused the inconsistencies in age-specific effects. Another suggestion of a confounding variable might be related to class norms. Previous research on class characteristics for example indicates that class normative beliefs about bullying can help explain bullying-related behaviours over and above individual characteristics (Pozzoli, Gini & Vieno, 2012; Salmivalli & Voeten, 2004). This does not take away that interventions that aim to promote agreeableness need to acknowledge the need to developmentally tailor their intervention efforts, as reflected in age- or grade-dependent curricula (Malti et al., 2016). This is also what we tried in our intervention program, as Appendix G illustrates.

4.1 Strengths and limitations

Due to the high tendency to report socially desirable with regard to prosocial and antisocial behaviour (Crothers & Levinson, 2004), one strength of our study was the use of peer nominations of bullying-related behaviour and prosocial and antisocial behaviour. By using peer nominations the chance of an error occurring due to a single reporter's experience reduces significantly. On the other hand, the use of peer nominations might have been disadvantageous since behavioural reputations sometimes tend to consolidate. This means that even though a child's behaviour may have changed over the year, their reputation for this behaviour persists with peers (Rubin, Bukowski & Parker, 2006). Therefore, on the positive side, the changes in the outcomes of this study can be considered changes in behaviour to an extent that can be observed by peers. In this way, we outperformed several other intervention programs that were unable to account for behavioural change. As Merrell et al. (2008) noted, the strongest effect sizes evaluating bullying prevention programs are generally found for how well participants know the intervention program and how they thought they would respond to bullying. On more direct measures of bullying-related behaviour, interventions generally had little positive effects, and sometimes even negative

effects. These negative effect can be explained by the phenomenon that children and teachers learn how to better recognize bullying-related behaviour through the intervention, and then recognizing the behaviour more often because of their increased awareness; i.e. the 'sensitization effect' (Frey, Hirschtein, Edstrom & Snell, 2005). This effect might partially explain the statistical non-significant effects of the intervention and delayed intervention on bullying.

A strength of our study was the use of three different intervention conditions combined with a longitudinal design. This first of all allowed us to compare a short and long duration of the intervention program. Second, it allowed us to follow-up the effects of the intervention program a year after the implementation. Third, in this way we could replicate the intervention condition with the delayed intervention condition. All three advantages of our design strengthen the internal validity of our outcomes. Additional strengths of the current undertaking are the use of multilevel growth curve modelling of hierarchical data, systematic implementation monitoring, and a relatively small amount of missing data combined with an attrition analysis. Moreover, because the statistical model for multilevel repeated measures data does not require the same number of measurement occasions per individual, all of the available data was incorporated into the analysis and weighted according to their presence. However, our study does show some of the methodological weaknesses mentioned by Baldry and Farrington (2007) concerning the evaluation of anti-bullying programs. A limitation of this study is the lack of a randomized design based on classes or schools. Despite the random assignment of parallel grades to the intervention or control group in the first year, the sample size was not big enough to assure comparability of children, teachers, and classes in the different conditions. This could have undermined the internal validity of our results; we cannot exclude that the effects found were due to other factors such as pre-existing differences between the trajectories of the four groups or the phenomenon of regression toward the mean. However, as said, the use of a longitudinal design with three different conditions strengthened the internal validity. Moreover, we attempted to account for pre-existing differences between the conditions by performing a growth curve analysis that takes the initial levels of the outcome variables into consideration. This suggests that the effects were not simply due to initial differences. In addition, 'contamination effects' could have occurred in the control group since all conditions were present in the same school. Future attempts to evaluate the effects a similar program should be

conducted by using different groups from different schools to avoid diffusion of treatment effects. Also, a bigger sample size would have strengthened our findings, since this would have resulted in a more appropriate comparison across conditions Moreover, the generalizability of our results is limited. Our sample of one school is not representative of Dutch schools, let alone schools in other countries. However, we do believe the outcomes of this study can be used in different contexts. Another limitation is our use of a single item for agreeableness during the first and second measurement occasion. Although this allowed us to have information on a lot of outcomes, it is possible that we may have lost a more complete and representative dimension of the construct of agreeableness. We tried to overcome this by using more items on agreeableness in the third and fourth measurement occasion. As described, the item-total correlation with for the item we used the first two occasions was relatively high.

4.2 Implications for policy and practice

Bullying in school is not only a great burden for teachers and classroom climate, but also for the perpetrators, victims and society, both financially and socially (Hawker & Boulton, 2000; Nansel et al., 2004; Nishina & Juvonen, 2005; Salmivalli & Isaacs, 2005; Soepboer, Veenstra & Verhulst, 2006). Therefore, there is an increasing need for education to prevent bullying-related behaviour. Despite its limitations, the current study has contributed to educational practice and literature on the effectiveness of anti-bully programs by investigating whether agreeableness might be a success-promoting factor of these type of interventions. The effects of our intervention program promoting agreeableness may be taken as promising venues in view of further improvement of class-based programs in education targeting antisocial and bullying-related behaviour.

The present findings may be helpful for administrators and school counsellors as they highlight the relevance of promoting agreeableness in school settings as a strategy for counteracting antisocial behaviour and bullying-related behaviour during middle childhood. Moreover, the results illustrate the importance of sustaining intervention programs aimed at the social development of children. Additionally, teachers interested in promoting positive developmental outcomes may find the intervention program as a useful tool to successfully build agreeableness in class. The intervention program provides insight into the environmental changes needed to promote agreeableness in a developmentally

appropriate way, which in turn could facilitate efforts to develop preventive actions at early stages in development. Because the seeds of anti- and prosocial behaviour most probably emerge in childhood (Eisenberg et al., 2002), knowledge about these years is most essential to understand and intervene in anti- and prosocial behaviour in education.

Chapter 6
Conclusion and Discussion

In this chapter we aim to recapitulate the four empirical studies in this dissertation, and to summarize and integrate their main findings and implications. The chapter starts with an overview of the issues that motivated this dissertation. This is followed by a summary of main findings derived from each of our empirical studies separately. We continue with a discussion of the integrative findings and contributions focussing on the overarching theoretical and methodological issues emerging from the empirical studies. Then, the integral limitations of our research are addressed. Finally, we discuss the implications for practice and further research that rise from our studies.

1. Research Motivation and Context

Exploring the role of moral components in anti- and prosocial behaviour in primary education is an issue of current scientific and practical interest. There is an increasing awareness of the need for educational systems to encourage the acquisition of prosocial values and behaviour and to discourage the acquisition of antisocial values and behaviour. However, the effects of intervening in anti- and prosocial behaviour in primary schools are mixed. Therefore, the starting point of our research concerned the question what underlies anti- and prosocial behaviour in middle childhood. Specifically, we argued that moral functioning might serve as a central process underlying children's anti- and prosocial behaviour in elementary school. The continuing decline in formerly coherent value systems and an increasing individualization in modern Western society make this even more relevant (Brown, Corrigan & Higgins-D'Alessandro, 2012; Fink & Slade, 2016; Rupp & Veugelers, 2003).

In our exploration, we drew on the theoretical framework of the Four Component Model of Rest (1983; 1986). This model currently offers the most adequate framework to examine the underlying psychological processes of moral behaviour: moral sensitivity, moral reasoning, moral motivation, and moral character. The first component, moral sensitivity concerns interpreting a situation in terms of how people's welfare is affected by possible actions of the subject. The second component, moral reasoning, regards integrating various considerations to determine what ought to be done. Moral motivation, the third component, concerns the importance people give to moral values (doing what is right) relative to other values (i.e. self-actualization). The fourth component is moral character and refers

to the ability to persist in a moral task in the face of obstacles (Rest, 1983; 1986; 1994).

We took different, but complementary paths in our exploration. In the first step of our research, assessment methods were developed and tested in the Dutch context of this study. Supportive evidence of reliability and validity of instrumentation clearly is a critical feature of meaningful research. In our second chapter we therefore elaborated on the assessment of emotions in the context of moral transgressions in the light of two important aspects of this particular assessment method, i.e. domain and developmental variability, and its links to important criterion measures, i.e. aggressive and prosocial tendencies, and sympathy. Moreover, its reliability and several aspects of its validity were tested. In the next two studies moral sensitivity, moral reasoning, moral motivation and moral character were examined in relation to the development of prosocial behaviour and a repetitive and intentional form of antisocial behaviour, bullying-related behaviour. Specifically, in the second study, the theory of marginal deviations of Caprara et al. (1992) was applied to explain development of prosocial behaviour. We examined whether initial marginal deviations (positive and negative) in prosocial behaviour lead to the development of prosocial behaviour when combined with the (accumulation of) moral components. In the third study relating the moral components to bullying-related behaviour the aim was to also take group characteristics into account. This is why we examined the relative contribution of the four components at both the individual and class level to five participant roles in the bullying context. In the fourth and last study the effects of an intervention stimulating children's behavioural tendencies towards agreeableness, an aspect of moral character, were analysed. Specifically, the aim was to investigate the effects of a class-based intervention program promoting agreeableness on bullying, assisting, defending, victimization, outsider behaviour, prosocial behaviour, and antisocial behaviour over the course of two school years. In this way the study contributed to the urgent need to know what is effective in moral education and the prevention of bullying.

2. Overview of main findings

In the first study we focused on the assessment of moral motivation, i.e. emotions following hypothetical transgressions in moral dilemmas. In Chapter 2 the

reliability and validity of the assessment of anticipated emotions in moral dilemmas was examined with a special interest in the domain and developmental specificity of the task. Although research on anticipated emotions in moral dilemmas has been published for decades, little work used statistical techniques such as factor analysis. This first attempt to look at the reliability and (aspects of) the validity of the instrument, helped to evaluate the custom for assessment of moral emotions. The instrument consisted of six scenarios covering three domains: fairness (not winning fairly, not keeping word), omission of a prosocial duty (not sharing, not helping) and victimization (verbal bullying, relational bullying). Anticipated emotions following the scenarios were coded as either negative (i.e. moral) or positive. The instrument appeared to be a reliable one-factor measure of anticipated emotions following hypothetical moral scenarios. The results further revealed some indications for the concurrent and predictive validity of the assessment of emotions in moral dilemmas. Interestingly, relations between anticipated emotions and prosocial tendencies were found in specific scenarios, namely the scenarios involving an omission of a prosocial duty (not sharing, not helping). This points to domain variability and to the importance of relating the situations of the moral scenarios to the behaviour of interest. Sympathy related to anticipated emotions following all scenarios. No relation was found between self-evaluated emotions and antisocial tendencies. Contrary to the expectations, emotions following hypothetical moral dilemmas did not show a strong developmental pattern.

Having identified the validity and reliability of the scale scores of the instrument representing moral motivation, we were more confident to look into the role of moral sensitivity, moral reasoning, moral motivation and moral character in (the development of) anti- and prosocial behaviour. Additionally, because anticipated emotions following moral dilemmas positively related to prosocial tendencies, but not to antisocial tendencies, we decided to study the role of the four moral components separately for prosocial and antisocial behaviour. Chapter 3 focussed on prosocial behaviour investigating the role of the (accumulation of) the moral components moral sensitivity, moral reasoning, moral motivation and moral character in predicting prosocial behaviour development. In Chapter 4 we switched our focus to the role of moral components in a specific form of antisocial behaviour, namely bullying behaviour. By studying the role of all moral components in both prosocial and antisocial behaviour, we tried to progress previous research empirically evaluating the Four Component Model of Rest (1983; 1986).

In Chapter 3 we studied whether the (accumulation of) moral components moderated the influence of initial marginal deviations (positive and negative) in prosocial behaviour on the development of prosocial behaviour over time. Marginal deviations in behaviour had been neglected in past research, which focused mainly on extreme groups or continuous dimensions of behaviour (Caprara, Dodge, Pastorelli & Zelli, 2007). The theory of marginal deviations argues that marginal deviations in behaviour also have the potential to develop into higher levels of this behaviour. Our study was among the first to study marginal deviations in prosocial behaviour, providing new insights in research and theorizing about prosocial behaviour development. We tested whether development in prosocial behaviour could be predicted by initial marginal deviations in prosocial behaviour. Moreover, based on the Four Component Model, the unique and aggregated effects of individual differences in moral functioning were considered to be relevant for prosocial behaviour development. Contrary to the theory of marginal deviations, the empirical findings of this study do not show that marginal deviations in prosocial behaviour have the potential to develop into higher or lower levels of prosocial behaviour over time. The development of initial prosocial behaviour into more pronounced prosocial behaviour only applied to marginally prosocial older children in the sample, and only in combination with high moral motivation. Moreover, marginally prosocial and marginally nonprosocial children tended to develop towards the mean of prosocial behaviour over time. Deviations from the norm of prosocial behaviour thus appeared to be compensated by behaviour in the opposite direction, instead of accumulating into stronger deviations from the norm of prosocial behaviour. This led us to speculate that the theory of marginal deviations mainly applies to behaviour with a high salience, such as aggressive or bullying behaviour. Additionally, moral sensitivity, moral reasoning, moral motivation and moral character did not significantly predict or moderate the development of prosocial behaviour. The relation between accumulation of moral components and the development of prosocial behaviour was stronger than the relation between the separate moral components and the development of prosocial behaviour. This preliminary finding could suggest that in order to develop prosocial behaviour, children might need to score high on all moral components instead of just one. This is in line with the Four Component Model assuming that all moral components must be in place in order to act morally.

In Chapter 4 we studied the role of the moral components moral sensitivity, moral reasoning, moral motivation and moral character in bullying-related behaviour. Bullying is a severe form of antisocial behaviour in which a perpetrator repetitively harms a victim. The (long-term) consequences for the victim, as well as the bully, are therefore not to be taken lightly. Doing justice to this form of antisocial behaviour we studied bullying as a group phenomenon. The participant roles of bully, assistant, defender, outsider and victim were compared with regard to moral sensitivity, moral reasoning, moral motivation, and moral character at both the individual and class level. Again we followed the theoretical framework of Rest and used a broader representation of measures of moral functioning than most studies do. Moreover, a more in-depth approach was taken for the operationalization of the four moral components by also aggregating them to the class level of the analysis. The results of this undertaking showed that all the moral components are related with the chance of taking a specific particular participant role in the bullying process. This supports an integrated analysis of several moral components for a deeper understanding of the moral precursors of bullying behaviour. Furthermore, it showed the importance of distinguishing between the different participant roles of bully, assistant, defender, victim and outsider. Also, the results demonstrated that class moral characteristics can help explain bullying-related behaviour over and above individual characteristics. Negative anticipated emotions following a moral transgression and low inhibitory control at the class level predicted high ratings of bullying. This suggests that targeting all children in class might be essential for successful interventions. Furthermore, lower conscientiousness was associated with greater chances of bullying (and even more in combination with lower inhibitory control), higher levels of moral reasoning were positively associated with bullying and assisting ,and higher sympathy and more negative anticipated emotions (and even more so when children were older) were positively associated with defending. Interestingly, the personality characteristic agreeableness was most (negatively) related to bullying and assisting behaviour when compared with defending and outsider behaviour. Stimulating agreeableness in middle childhood therefore seems a good place to start in order to prevent bullying at school. Accordingly, this study has gone some way towards enhancing our understanding of the precursors of children's behaviour in bully situations.

Since our research showed that agreeableness was most (negatively) associated with bullying-related behaviour and other research already found similar results, we argued that agreeableness might be a success-promoting factor of interventions targeting bullying-related behaviour and antisocial behaviour in middle childhood. Interestingly, agreeableness is also assumed to be the most malleable of the personality dimensions and most susceptible to change in light of environmental inputs. Therefore, the aim of Chapter 5 was to examine the effectiveness of an educational intervention on children's behavioural tendencies towards agreeableness in order to target bullying-related behaviour and anti- and prosocial behaviour. By solely focussing on agreeableness as the potential underlying success-promoting factor of the intervention, information was obtained about the explaining process of the effectiveness of an intervention (Does stimulating agreeableness lead to a reduction of antisocial behaviour and an increase in prosocial behaviour?) as well as evidence of effectiveness (Does the intervention reduce antisocial behaviour and increase prosocial behaviour?). Specifically, this chapter investigated the effects of a class-based intervention program promoting agreeableness in middle childhood on bullying-related behaviour, and anti- and prosocial behaviour. These effects were examined by comparing children in a control condition with three intervention conditions at the beginning and at the end of two school years: (a) the intervention condition in which the intervention was implemented during the first schoolyear, (b) the delayed intervention condition in which the intervention was implemented during the second schoolyear, and (c) the extended intervention condition in which the intervention was implemented during both school years. An encouraging and important result is that we succeeded in the stimulation of agreeableness in all intervention conditions. As expected, promising effects of the class-based intervention were also found for all conditions for reducing bullying-related behaviour and antisocial behaviour. Compared to the control condition, the children in the intervention conditions showed a decrease in antisocial behaviour, bullying, assisting, and victimization and an increase in outsider behaviour over time. These positive effects were especially visible in the extended intervention condition. Moreover, a striking number of gender and age specific effects were found. The role of agreeableness in predicting the other outcomes of the intervention was relatively small. Therefore, it is not sure whether the changes in agreeableness caused changes in the other outcomes. However, there are some

indications that stimulating agreeableness in primary education has the potential to redirect negative trajectories of functioning. Overall, findings suggest that promoting agreeableness may serve to counteract antisocial behaviour.

3. Integrative findings and contributions

In our exploration of the role of moral components in anti- and prosocial behaviour in primary education we came across several integrative findings. Before discussing these findings we would like to stress that both dimensions of our exploration - moral components and anti- and prosocial behaviour - are broad constructs. Therefore, our choices with regard to their operationalization and assessment directly and indirectly steer our findings. Correspondingly, the results of the previous chapters as well as the integrative findings we present here should be interpreted in the light of these choices. Not only are both dimensions of our exploration - moral components and anti- and prosocial behaviour - broad constructs, they also relate to one another in different ways. However, taking everything into account, our research led to the insight that moral functioning can be an important process leading to anti- and prosocial behaviour in middle childhood. Therefore, the main contribution of our research is the identification of potential success-promoting moral factors for intervening in anti- and prosocial behaviour in primary education. Additionally, seven integrative findings emerge when deepening our main contribution.

Our first integrative finding is that not all aspects of moral functioning, i.e. not all moral components, were equally related to both anti- and prosocial behaviour. In order to illustrate this, we will briefly discuss the separate moral components and their relations with anti- and prosocial behaviour in our studies. The strongest relations were found between *moral character*, operationalized as agreeableness, conscientiousness, and inhibitory control, and both anti- and prosocial behaviour. Agreeableness was negatively related to the roles of bully and assistant in comparison to the defender and outsider roles and a negative association appeared between conscientiousness and bullying. This negative relation was even strengthened when children inhibitory control was lower (Chapter 4). Additionally, the relations between *moral motivation* and anti- and prosocial behaviour were weak. Moral motivation was positively related to prosocial behaviour, but not to antisocial behaviour (Chapter 2) and prosocial

behaviour development (Chapter 3), defending, bullying and assisting (Chapter 4). The relations between *moral reasoning* and anti- and prosocial behaviour were opposite our expectations. Moral reasoning, did not significantly predict prosocial behaviour development (Chapter 3), but positively related to bullying and assisting and negatively to being victimized (Chapter 4). The relations between *moral sensitivity*, operationalized as sympathy, and anti- and prosocial behaviour were merely inconsistent. Moral sensitivity did not significantly predict prosocial behaviour development (Chapter 3) nor bullying or assisting, but did relate to defending behaviour in bullying situations (Chapter 4).

The second integrative finding is that the predictive value of the four moral components differs for prosocial behaviour and antisocial behaviour (e.g. Baumeister et al., 2001; Rothbart & Park, 1986; Krueger, Hicks & McGue, 2001; Malti & Krettenauer, 2013). For example, a positive relation was found between moral motivation and prosocial tendencies, but no relation between moral motivation and antisocial tendencies (Chapter 2). Another illustration of this integrative finding is that the moral components related differently to the different participant roles in the bullying process (Chapter 4). Defending, a clear prosocial act, was significantly related to sympathy, whereas assisting and bullying, clear antisocial acts, were not. Furthermore, the intervention promoting agreeableness did not have the same effects on pro- and antisocial behaviour (Chapter 5). Illustrative is that the children in all intervention conditions showed a decline in antisocial behaviour, while their prosocial behaviour decreased. This again indicates that different processes underlie pro- and antisocial behaviour.

The third integrative finding is that the combination or accumulation of moral components is predictive for pro- and antisocial behaviour. We not only examined the relation between anti- and prosocial behaviour and the individual four moral components, we also investigated whether the accumulation or combination of moral components might be related to anti- and prosocial behaviour. This is in line with Rest's (1983; 1986) theory behind the Four Component Model. Since he clearly states that all moral components are needed in order to behave in a moral manner, the interaction effects between moral components were incorporated in predicting bullying-related behaviour (Chapter 4) and we hypothesised that the accumulation of moral components might predict prosocial behaviour development (Chapter 3). Two interaction effects were found. First, sympathy was found to relate to assisting behaviour in bullying situations,

but only in combination with inhibitory control. Second, the relation between assisting and agreeableness became stronger with higher levels of sympathy (Chapter 4). Moreover, the positive linear and negative quadratic relation between the accumulation of moral components on prosocial behaviour development was stronger than the relations between the separate moral components and prosocial behaviour development (Chapter 3). This preliminary finding suggests that in order to develop prosocial behaviour it is beneficial to score high on more than one moral component.

A fourth integrative finding is that the relation between moral components and anti- and prosocial behaviour differed between individual and class moral components. Interestingly, class moral emotions and class inhibitory control were related to bullying-related behaviour. At the individual level, however, these relations were not present (Chapter 4). This indicates that class moral characteristics can help explain bullying-related behaviours over and above individual moral characteristics (Gini et al., 2014; 2015; Pozzoli, Gini & Vieno, 2012). More generally, group processes seem to play an important role in the relations between moral functioning and anti- and prosocial behaviour. This might also explain the phenomenon of compensation that was found with regard to prosocial behaviour development (Chapter 3). In this chapter, children with initial deviations in prosocial behaviour appeared to develop towards the mean of prosocial behaviour over time. This supports the finding of Dodge (2006) that deviations from the norm, i.e. the mean, tend to be compensated by actions in the opposite direction. This tendency to develop to the norm might be due to processes of social influence among classmates. It is known that groups help define the type and range of relationships and interactions that are likely or permissible (e.g. Rubin, Bukowski & Parker, 2006).

The fifth integrative finding emerging from our research is that the associations between moral components and anti- and prosocial behaviour sometimes also depends on individual characteristics such as gender and age. The most remarkable finding related to gender is that the intervention program stimulating agreeableness showed differential gender effects (Chapter 5). Boys showed a stronger decline in antisocial behaviour, bullying, and assisting than girls in the intervention and extended intervention condition. Additionally, age seemed to play an important role in the relations between moral motivation and anti- and prosocial behaviour. Interaction effects were found for the links between age and

moral motivation when predicting antisocial behaviour, prosocial behaviour and defending (Chapter 2, 3 and 4). Overall, however, the effects in our studies transcended both gender and age. This indicates stable inter-individual differences in the relation between moral components and anti- and prosocial behaviour (Arsenio, Gold & Adams, 2004; Krettenauer & Eichler, 2006; Malti & Ongley, 2014).

Our sixth integrative finding is that the cross-sectional relations between the four moral components and anti- and prosocial behaviour differed from longitudinal relations. We found a cross-sectional positive relation between moral motivation and prosocial behaviour (Chapter 2), but, the positive relation between moral motivation and the development of prosocial behaviour was only found for younger nonprosocial children (Chapter 3). This led us to believe that the development of anti- and prosocial behaviour might perhaps require extra or other success-promoting moral factors than anti- and prosocial behaviour at one time point.

Our last integrative finding is that we obtained information about the process behind the effectiveness of intervening in anti- and prosocial behaviour. To this end we not only identified success-promoting factors of intervention programs aimed at anti- and prosocial behaviour (Chapter 3 and 4), we also developed an intervention program to isolate the effects of promoting agreeableness on the reduction of antisocial behaviour and promotion of prosocial behaviour (Chapter 5). Thus, instead of solely focussing on the evidence of effectiveness of a program (Does the intervention reduce antisocial behaviour and increase prosocial behaviour?) we also gained insight into the processes explaining the effectiveness of the intervention (Does stimulating agreeableness lead to a reduction of antisocial behaviour and an increase in prosocial behaviour?). The results of this undertaking showed that agreeableness appeared to be a malleable construct (Bergeman et al., 1993; Graziano & Eisenberg, 1997), and potentially negatively affected bullying-related and antisocial behaviour in middle childhood. In this way, our research contributed to the existing literature by focussing on the process of how an intervention might work (e.g. Gravemeijer & Kirschner, 2007).

4. Limitations and strengths

Despite its strengths and novel focus on relations between several dimensions of moral functioning and different forms of both pro- and antisocial

behaviour, this dissertation also has several limitations as a whole. In the following these limitations will be discussed in conjunction with the strengths of the dissertation. We start by pointing out that the choices we made with regard to the use of the Four Component Model as well as our operationalizations of the four moral components are open for debate. Second, we discuss the limitations and strengths related to our way of assessing the constructs under study. Then we elaborate on the limitations and strengths concerning the design of our studies. Fourth, we consider the strengths of the chosen analyses.

First of all, the choices we made related to the use of the Four Component Model and the operationalizations of the constructs under study could be considered a limitation. There are indications that the Four Component Model is outdated after a shift from more cognitive to affective and neuro/biological paradigms (e.g. Narvaez & Vaydich, 2008; Turiel; 1983; 1998; 2006). Another downside of the model is that it has mainly been used in adolescents and adults and therefore lacks empirical support in younger samples (Rest, 1999). Moreover, there are other relevant constructs that are not in the Four Component Model that explain anti- and prosocial behaviour. For example, the Four Component Model does not entail the construct of moral disengagement, while this is nowadays considered an important fundamental mechanism explaining antisocial behaviour (Bandura et al., 1996; Gibbs, 2014). Nonetheless, the combination of the comprehensive and broad account of moral processes in the Four Component Model and the emphasis on the complicated interaction between these moral processes provided us with a valuable framework to study the underlying processes of anti- and prosocial behaviour. Also, our research provided some empirical evidence for the legitimacy of the model. Remains to be said that the operationalizations of the constructs under study could be optimized. For example, our operationalization of the construct of moral sensitivity was limited to sympathy, its more affective aspect. Unfortunately our attempts to capture the whole scope of moral sensitivity were unsuccessful, leading to sympathy being the best representation of moral sensitivity we could find in children.

A second limitation of our research concerns the assessment of the constructs in our dissertation. The assessment these constructs was limited in time and capacity, leaving room for improvement. For instance, the behavioural constructs in our studies were assessed with peer nominations. Peer nominations have three disadvantages: behavioural reputations sometimes consolidate, a child's

judgment of a peer might be influenced by his/her own abilities and behaviour, and they are generally gender-biased (Rubin, Bukowski & Parker, 2006; Warden, Cheyne, Christie, Fitzpatrick & Reid, 2003; Warden & Mackinnon, 2003). However, besides these three downsides, peer nominations are a highly valued way to assess behavioural constructs; the chance of an error occurring due to a single reporter's experience reduces significantly and peers can identify characteristics and relationships of children that are considered relevant from their perspective (Rubin, Bukowski & Parker, 2006). Moreover, the changes in the behavioural outcomes of this study can be considered changes in behaviour to an extent that can be observed by peers (Merrell et al., 2008). A second illustration of possible improvement of our assessment methods concerns the assessment of agreeableness and conscientiousness. This assessment exclusively relied on teacher-reports of a single item during the first and second measurement occasion. Although this allowed us to have information on a lot of outcomes, we may have lost a more complete and representative dimension of these constructs. Yet, we tried to overcome this limitation by using more items on agreeableness in the third and fourth measurement occasion. A third example of the limitations in assessment relates to the relative low Cronbach's Alpha for the scale scores of moral reasoning. Therefore, the results with regard to this measure must be interpreted with care. Also, the measure of moral emotions might have benefited from the assessment of intensity of the emotions instead of their valence (Malti & Krettenauer, 2013). A last remark with regard to assessment is that the data in our study was based on self-, other- and teacher ratings, and may therefore be subject to social desirability. Observational measures would be a welcome addition to this type of research. However, combining peer-, teacher- and self-reports already added to the credibility of our results. Moreover, the chosen assessment methods allowed us to empirically assess both pro- and antisocial behaviour, bullying-related behaviour, and the entire Four Component Model.

A third limitation of this dissertation lies in our research design. Despite the many advantages and the suitability of this design, it also has two main drawbacks. First, the children participating in the current inquiry were not a random sample, since all schools in our sample were located in the Northern part of the Netherlands resulting in a homogeneous sample comprised of predominantly white pupils of Dutch descent (97.5%). Hence, it remains unclear whether the findings in our empirical studies can be generalized to other settings. However,

great care was taken in assuring variability in school denomination (three catholic, two protestant, six public), size (from 37 to 307 children, M=119.2 and SD=71.9), location (seven in rural areas and five in cities), and mixing of grades (five with single-graded classroom, six with multi-graded classrooms). Still, the conclusions derived in our Chapters are limited to the children participating in our studies. Replication with another sample would overcome this limitation. Second, our research design did not capture all variability in behaviour and moral functioning. It remains unclear whether the data collected at the four particular occasions in two years were representative of what the children were exposed to in the course of these years. Also, we did not look into specific interaction processes. For example, the causal mechanisms underlying the relation between marginal deviations in prosocial behaviour, the (accumulation of) moral components and prosocial behaviour development remain unclear. Though, our quasi-experimental study did contribute to making causal inferences relating to the relation between agreeableness and anti- and prosocial behaviour. A strength of our study was the use of three different intervention conditions combined with a longitudinal design. This first of all allowed us to compare a short and long duration of the intervention program. Second, it allowed us to follow-up the effects of the intervention program a year after the implementation. Third, in this way the intervention condition was replicated with the delayed intervention condition. All three advantages of our design strengthen the internal validity of our outcomes. Moreover, the design of this research project had several other advantages that enabled us to investigate the moral processes behind behavioural change in the setting of primary education. The fact that it was a longitudinal study allowed us to examine the role of moral components in predicting pro- and antisocial behaviour. This study is one of the few field studies on the Four Component Model with more than one measurement occasion (see also Hardy, 2006; Morton, Worthley, Testerman & Mahoney, 2006). Another advantage of our research design is that whole school classrooms were targeted so that a complete social environment with a meaningful boundary could be investigated. More generally, the set-up of the study allowed us to look into moral functioning and behaviour in a natural and therefore meaningful setting using a wide age range. Also, the study had a very high response rate of varying from 97.3% to 99% for the different measurement occasions. This enabled us to get a complete picture of the whole scope of moral processes and anti- and prosocial behaviours going on within a classroom context over time (Neal, 2008).

A strength of this dissertation lies in its analyses, and specifically in the use of multilevel regression analysis. An advantage of the use of this multilevel approach is that the school, class, individual and/or time level are distinguished. This recognition of hierarchical structures creates a more correct estimation of standard errors and the tests of specific effects for single dependent variables are also more powerful (Snijders & Bosker, 2000). Further, a multilevel regression analysis allows for investigation of group phenomena, which appeared to be particularly valuable (see Chapter 4). Finally, since the statistical model for multilevel repeated measures data does not require the same number of measurement occasions per individual, all of the available data could be incorporated into the analysis and weighted according to their presence (see Chapter 5).

5. Implications for further research and practice

Based on our studies we would like to formulate lines of implications for research and practice. We start by formulating four implications for further research and continue with three implications for practice.

5.1 Implications for further research

One important implication for research is that it is worth striving to optimize the reliability and validity of the instrumentation used in this dissertation. Even though supportive evidence of the reliability and validity of the assessment of moral motivation was found, it is crucial that assessment tools continue to be optimized and tested. Moreover, future studies are encouraged to report about the structure and reliability of the scale score of the instrument assessing anticipated emotions to confirm and complement our findings. The same applies to the optimization of the operationalization and assessment of the other moral components that was not extensively reported in this thesis. For example, future research might try to optimize the operationalization and assessment of the construct of moral sensitivity, since our assessment of sympathy did not fully capture this construct. Also, future studies might benefit from observational data. For example, research has shown that peers, especially girls, are more likely to nominate girls as being prosocial and nominate boys as being bullies (Warden, Cheyne, Christie, Fitzpatrick & Reid, 2003; Warden & Mackinnon, 2003). Fewer

differences have been found in observational studies (Fabes, Martin & Hanish, 2002; Zahn-Waxler et al. 2001). Also, future studies could benefit from more advanced measures of agreeableness and conscientiousness from both the teacher's and child's perspective and a more reliable measure of moral reasoning. These adjustments in instrumentation will yield important additional knowledge on the relations between the (accumulation of) individual and class moral components and (in the development of) pro- and antisocial behaviour. More knowledge in the areas of moral developmental research has significant implications for the design of preventive interventions aimed at increasing care, and social justice in children (Malti & Ongley, 2014).

Second, to fully consider the influence of moral functioning on change in anti- and prosocial behaviour, further research should use longitudinal and fully experimental designs that deepen the understanding of how moral precursors influence anti- and prosocial behaviour. Even though we made use of longitudinal data, a lot of change could have happened within, before, or after the four time points in our design. Furthermore, it remains unclear whether the data at these time points is representative of what the children were exposed to in the course of their school years. Future studies are recommended to examine the reactions of peers and teachers on children behaviour and their effect on future behaviour in order to test the claims underlying the development of anti- and prosocial behaviour.

Third, future studies are recommended to take the class level and group processes into account when studying the associations between moral functioning and anti- and prosocial behaviour. Our research indicates that peer relations play an important role in the relations between moral functioning and anti- and prosocial behaviour. Additionally, it would be interesting to look at the links between the moral components and anti- and prosocial behaviour using a social network perspective (e.g. Huitsing, 2014). Social network analysis enables researchers to model a variety of dependence patterns, such as dyadic and group level processes (Cook, 2011). Moreover, in social network analysis participant roles (i.e. bully, assistant, defender, victim, and outsider) no longer have to be regarded as 'fixed', because it allows investigation of the variation in children's behaviour towards different classmates (Huitsing & Veenstra, 2012). In this way, more nuanced information can be obtained about the children's social status.

Last, studying success-promoting factors and underlying processes of interventions targeting bullying-related behaviour in middle childhood remains an

important topic for future research. In our undertakings an intervention program was developed to isolate the effects of promoting agreeableness on the reduction of antisocial behaviour and promotion of prosocial behaviour. Future attempts to evaluate the effects of an intervention program directed at the promotion of agreeableness should be conducted using a randomized design at class- or school level. This would strengthen causal claims behind the effects of promoting agreeableness on the reduction of antisocial behaviour and promotion of prosocial behaviour. Also, this would strengthen our idea that stimulating agreeableness in primary education has the potential to redirect negative trajectories of functioning. Future research could also look into the influence of other moral components, such as moral motivation, to disentangle the (causal) effects of different moral components on the development of anti- and prosocial behaviour. A related avenue for future research is to compare the effects of the stimulation of several moral components on anti- and prosocial behaviour. As the accumulation of moral components was predictive of both pro- and antisocial behaviour, stimulating multiple components in an intervention could be more beneficial than focusing on only one moral component. A future study could for example compare four different intervention conditions: one focused on the promotion of agreeableness, one focused on the promotion of sympathy, one on both, and a control condition. Moreover, it would be interesting to compare the effects of promoting two moral components to the effects of promoting more than two moral components, or, in line with the Four Component Model, all four moral components.

5.2 Implications for practice

The main practical implication emerging from our research relates to the potential of the intervention program promoting agreeableness. This intervention program may be taken as a promising venue in view of further improvement of class-based programs in education targeting antisocial and bullying-related behaviour. The results of our research may be helpful for both administrators and school counsellors as they highlight the potential of promoting agreeableness in school settings as a strategy for counteracting aggressive tendencies and bullying-related behaviour during middle childhood. Additionally, teachers interested in promoting positive developmental outcomes may find the intervention program a useful tool to successfully build agreeableness in class and thereby discourage antisocial behaviour. The intervention program provides insight into the

environmental changes needed to promote agreeableness in a developmentally appropriate way, which in turn could facilitate efforts to develop preventive actions at early stages in development. Because the seeds of anti- and prosocial behaviour emerge in childhood (Hepach, Vaish, Grossmann & Tomasello, 2016; Malti & Dys, in press), intervening in these years is most essential to counteract antisocial behaviour.

Additionally, the intervention program aimed at the promotion of agreeableness might benefit from the supplementation of the promotion of other moral components. As the accumulation of moral components is predictive of both pro- and antisocial behaviour, stimulating multiple moral components in an intervention could be more beneficial than focusing on one moral component. Also, intervention programs aimed at affecting bullying might profit from enhancing all children's inhibitory control and moral emotions since class inhibitory control and moral emotions were found to be related to bullying behaviour. Though, as previously recommended, the effectiveness of such undertakings first needs to be established. More generally, targeting the whole class and not only the individual bully and/or victim seems a promising venue for intervening in bullying-related behaviour in education. Furthermore, practitioners should be aware that counteracting antisocial behaviour requires other actions than promoting prosocial behaviour. We argued that prosocial behaviour seems to develop according to compensation, whereas aggression develops according to amplification. Stimulating prosocial behaviour might therefore require a norm shift of the whole classroom instead of individual encouragements.

Additionally, our findings suggest that the effectiveness of interventions targeting agreeableness grows when carried out over a longer period of time. The positive effects of the intervention were especially visible in the extended intervention condition. Thus, changing behavioural tendencies in a class environment might require patience and persistence. Both policymakers and teachers would therefore be wise to persevere in influencing anti- and prosocial behaviour in education. Unfortunately, however, most efforts to improve the social environments of schools are not sustained (Supovitz & Weinbaum, 2008). We think these ephemeral efforts should change in order to contribute to the social outcomes of education.

6. Concluding remarks

Altogether, our research led to the insight that moral functioning can be an important process leading to anti- and prosocial behaviour in middle childhood. Therefore, the main contribution of our research is the identification of potential success-promoting moral factors for intervening in anti- and prosocial behaviour in education. Our findings also indicate that not all aspects of moral functioning, i.e. not all moral components, were equally related to anti- and prosocial behaviour. The strongest relations were found between moral character and anti- and prosocial behaviour. Our intervention program aimed at the promotion of agreeableness also stressed this finding. Additionally, the accumulation and combination of moral components also contributed to anti- and prosocial behaviour on top of the individual contributions of the moral components. In a related fashion, group processes seem to play an important role in the relations between moral functioning and anti- and prosocial behaviour next to individual processes. However, the aforementioned processes had a different contribution to prosocial compared to antisocial behaviour. This means that counteracting antisocial behaviour in education requires another focus than the promotion of prosocial behaviour. Further advancements in the understanding of how individual and class moral components affects anti- and prosocial behaviour may inform educational researchers and practitioners alike about the promotion of prosocial behaviour and targeting antisocial behaviour.

Appendices

Appendix A

Matching cartoons for the girl version of the not sharing scenario by Emma Wilson

Appendix B

Coding scheme attributed emotions in moral transgressions

Main Categories	Examples
(1) **Happy**	**happy, good, great, proud, pleased, glad, fine, joyful, excited, satisfied, nice, enthusiastic, helpful,** not bad, funny.
(2) **Neutral**	**normal, okay, as usual, regular,** wouldn't care, wouldn't feel bad, wouldn't be affected, not concerned, wouldn't think about it.
(3) **Angry**	**angry, mad, frustrated, irritated, annoyed, furious, rage, defensive, offended.**
(4) **Scared**	**scared, afraid, frightened, horrified, terrified, anxious, worried, nervous.**
(5) **Bad**	**bad, upset, terrible, miserable, lousy,** unhappy, not good, not great, not nice, not helpful, ungrateful, rude, wouldn't feel proud, wouldn't feel right, uneasy, wrong.
(6) **Sad**	**sad, sorry, sorrow**.
(7) **Guilty**	**guilty, regretful, remorseful, blameworthy.**
(8) **Ashamed**	**embarrassed, ashamed, shameful, disgraced, humiliated.**
(9) **Disgusted/mean**	**disgusted, mean, selfish, greedy, unfair, disappointed, gross, sick.**

Appendix C

Mean and standard deviation of age, prosocial tendencies, antisocial tendencies, and sympathy for positive and negative attributed emotions at the same time and a year later for all the scenarios separately and across scenarios

Scenario	Anticipated emotions	*M* (*SD*) Age	T1: *M* (*SD*) Antisocial tendencies	T1: *M* (*SD*) Prosocial tendencies	T2: *M* (*SD*) Antisocial tendencies	T2: *M* (*SD*) Prosocial tendencies	T1: *M* (*SD*) Sympathy	T2: *M* (*SD*) Sympathy
Not sharing	Positive	8.82 (1.9)	16.5 (21.2)	44.2 (20.1)	15.5 (20.5)	42.1 (20.0)	2.00 (.49)	**1.88 (.51)**
	Negative	9.14 (1.8)	17.7 (20.8)	46.9 (18.7)	16.2 (20.2)	47.4 (20.5)	2.11 (.48)	**2.07 (.50)**
Not helping	Positive	**8.43 (1.8)**	17.4 (21.9)	**40.5 (19.4)**	16.2 (20.6)	**42.9 (20.5)**	**1.91 (.50)**	**1.91 (.51)**
	Negative	**9.23 (1.8)**	17.5 (20.6)	**47.6 (18.6)**	15.9 (19.9)	**47.5 (20.4)**	**2.13 (.46)**	**2.07 (.50)**
Not winning fairly	Positive	9.02 (1.8)	16.1 (20.5)	45.8 (19.6)	14.6 (20.7)	47.0 (20.7)	**2.00 (.48)**	**1.96 (.50)**
	Negative	9.14 (1.8)	18.0 (20.9)	46.8 (20.5)	16.5 (20.5)	46.7 (20.5)	**2.13 (.48)**	**2.07 (.50)**
Not keeping word	Positive	**8.90 (1.8)**	17.6 (21.4)	**43.1 (19.1)**	14.9 (19.3)	**44.0 (20.2)**	**2.00 (.49)**	**1.96 (.51)**
	Negative	**9.18 (1.8)**	17.5 (20.7)	**47.7 (18.7)**	16.4 (20.4)	**47.7 (20.6)**	**2.13 (.47)**	**2.07 (.50)**
Relational bullying	Positive	**8.51 (1.8)**	17.2 (21.6)	**42.4 (19.5)**	16.1 (21.7)	**43.6 (21.0)**	**1.97 (.51)**	2.00 (.54)
	Negative	**9.26 (1.8)**	17.8 (20.8)	**47.6 (18.7)**	16.0 (19.8)	**47.5 (20.3)**	**2.13 (.47)**	2.06 (.49)
Verbal bullying	Positive	9.13 (1.9)	16.4 (20.1)	45.2 (19.4)	15.0 (20.4)	45.5 (20.4)	**2.03 (.50)**	**1.97 (.51)**
	Negative	9.10 (1.8)	17.9 (20.9)	46.9 (18.8)	16.3 (20.5)	47.1 (20.5)	**2.11 (.48)**	**2.06 (.50)**
Total		.11*	.02	.11*	.03	.08*	.16*	.13*

Notes. In bold: significant differences using $p<0.05$.

* Pearson correlation coefficients (r) is significant using $p<0.05$.

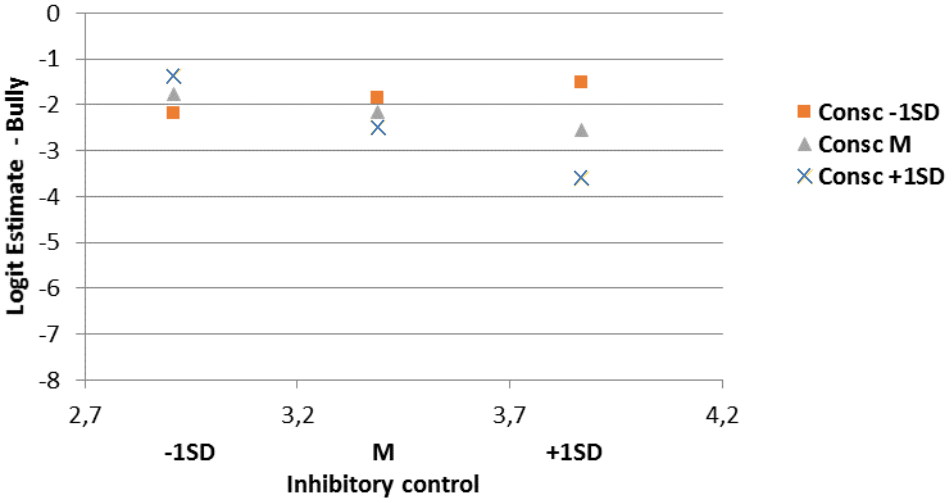
Appendix D

Multilevel regression analysis predicting the development of prosocial behaviour (n=491)

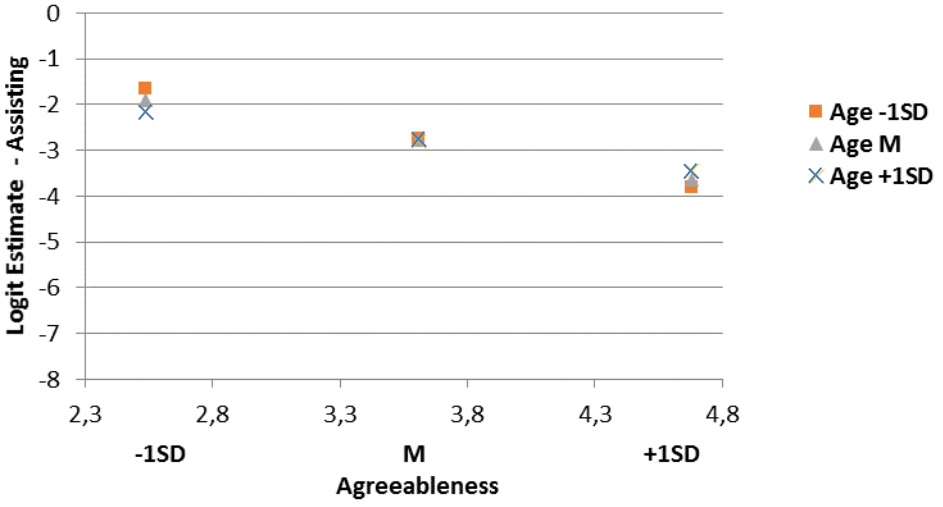
	Central moral processes model**		Accumulation model**	
Fixed effects	*Estimate*	*S.E.*	*Estimate*	*S.E.*
Intercept	-.628	2.254	-.899	2.188
Level 1 (individual) variables				
Girl	6.046*	2.344	5.929*	2.188
Age	1.649*	.746	1.933*	.668
Socioeconomic status	.888	.791	.647	.785
Scholastic ability	2.381*	.664	2.180*	.678
Initial prosocial behaviour	-.392*	.084	-.423*	.085
Marginal deviations in prosocial behaviour	.182	2.115	1.158	1.996
Sympathy	2.108	2.461		
Moral reasoning	-2.994	7.820		
Moral emotions	1.959	5.105		
Agreeableness	.046	1.291		
Conscientiousness	-.314	1.210		
Inhibitory control	-3.322	2.262		
Accumulation score			4.244	2.834
Accumulation score^2			-1.063	.768
Sympathy*Marginal deviations	2.779	3.341		
Moral reasoning*Marginal deviations	-6.655	9.615		
Moral emotions *Marginal deviations	4.131	6.796		
Agreeableness*Marginal deviations	.099	1.659		
Conscientiousness*Marginal deviations	-1.064	1.682		
Inhibitory control *Marginal deviations	-2.629	3.079		
Accumulation score* Marginal deviations			1.278	5.741
Accumulation score^2* Marginal deviations			-.754	1.575
Marginal deviations*Girl	.320	2.442	-.924	2.224
Marginal deviations*Age	-.065	.618	-.111	.600
Sympathy*Girl	-.509	3.259		
Moral reasoning*Girl	4.883	9.735		
Moral emotions*Girl	2.569	6.026		
Agreeableness*Girl	.488	1.710		

Conscientiousness*Girl	-1.804	1.587		
Inhibitory control*Girl	-.683	3.246		
Accumulation score*Girl			1.294	3.763
Accumulation score^2*Girl			-.319	.987
Sympathy*Age	-.366	.876		
Moral reasoning*Age	1.527	2.768		
Moral emotions*Age	4.491*	1.665		
Agreeableness*Age	.870	.487		
Conscientiousness*Age	.296	.451		
Inhibitory control*Age	1.466	.853		
Accumulation score*Age			-.755	.983
Accumulation score^2*Age			.253	.262
Sympathy*Marginal deviations*Girl	.163	4.769		
Moral reasoning*Marginal deviations*Girl	6.611	12.839		
Moral emotions *Marginal deviations*Girl	1.171	8.298		
Agreeableness*Marginal deviations*Girl	1.437	2.333		
Conscientiousness*Marginal deviations*Girl	-.438	2.307		
Inhibitory control *Marginal deviations*Girl	1.018	4.407		
Accumulation score * Marginal deviations*Girl			3.364	7.715
Accumulation score^2* Marginal deviations*Girl			-.039	2.027
Sympathy*Marginal deviations*Age	-.245	1.236		
Moral reasoning*Marginal deviations*Age	2.789	3.472		
Moral emotions *Marginal deviations*Age	-6.752*	2.158		
Agreeableness*Marginal deviations*Age	1.173	.644		
Conscientiousness*Marginal deviations*Age	1.005	.637		
Inhibitory control *Marginal deviations*Age	.438	1.177		
Accumulation score* Marginal deviations*Age			1.353	1.986
Accumulation score^2* Marginal deviations*Age			.126	.531
Random effects	*Var. Comp.*	*S.E.*	*Var. Comp.*	*S.E.*
School level variance	0.000	0.000	0.000	0.000
Class level variance	118.494	33.577	125.526	35.594
- Slope for Girl	54.651	21.832	50.830	21.416
- Slope for Scholastic ability	3.950	2.896	4.856	3.284
- Covariance Girl	-41.371	21.702	-42.781	22.243
- Covariance Scholastic ability	-.477	7.023	-.721	7.689
- Covariance Girl Scholastic ability	-13.779	6.050	-10.861	6.144
Individual level variance	114.665	8.080	119.592	8.461

Appendix E

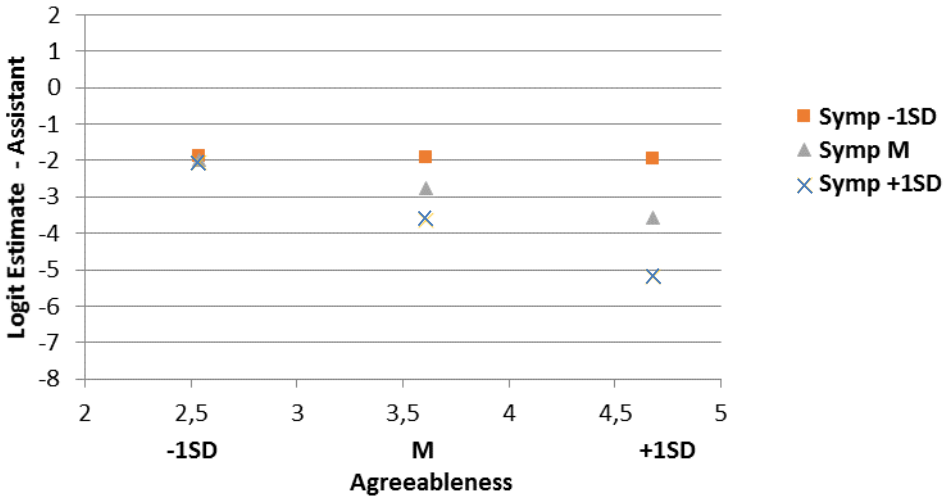


Interaction effect of conscientiousness on the effect of inhibitory control predicting the participant role of bully versus outsider



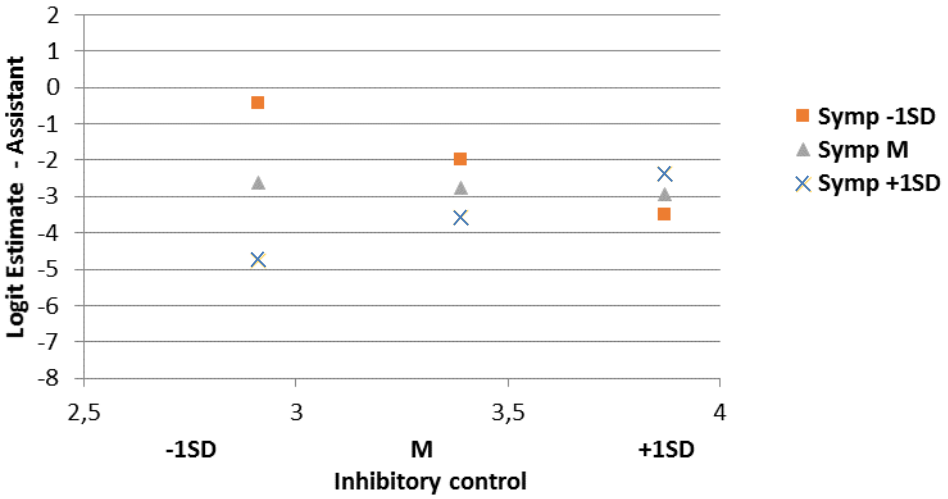
Interaction effect of age on the effect of agreeableness predicting the participant role of assistant versus outsider

The multinomial logit estimate for someone who scored 1SD below the mean of age and agreeableness was -1.66 ((-2.77+(-.03*-1)+(-.86*-1)+(.22*-1*-1)), corresponding OR=0.19); the multinomial logit estimate for someone who scored 1SD below the mean of age and 1SD above the mean of agreeableness was -3.82 (OR=0.02); the multinomial logit estimate for someone who scored 1SD above the mean of age and 1SD below the mean of agreeableness was -2.16 (OR=0.12); and the multinomial logit estimate for someone who scored 1SD above the mean of age and agreeableness was -3.44 (OR=0.03).



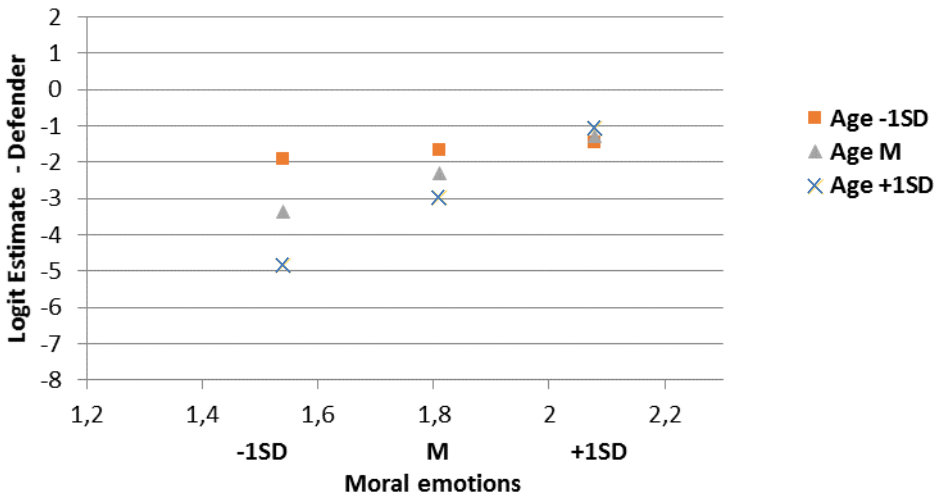
Interaction effect of sympathy on the effect of agreeableness predicting the participant role of assistant versus outsider

The multinomial logit estimate for someone who scored 1SD below the mean of sympathy and agreeableness was -1.87 ((-2.77+(-.80*-1)+(-.86*-1)+(-.76 *-1*-1)), corresponding OR=0.15); the multinomial logit estimate for someone who scored 1SD below the mean of sympathy and 1SD above the mean of agreeableness was -2.07 (OR=0.13); the multinomial logit estimate for someone who scored 1SD above the mean of sympathy and 1SD below the mean of agreeableness was -1.95 (OR=5.30); and the multinomial logit estimate for someone who scored 1SD above the mean of sympathy and agreeableness was -5.19 (OR=0.01).



Interaction effect of sympathy on the effect of inhibitory control predicting the participant role of assistant versus outsider

The multinomial logit estimate for someone who scored 1SD below the mean of sympathy and inhibitory control was -0.44 ((-2.77+(-.80*-1)+(-.16*-1)+(1.37*-1*-1)), corresponding OR=0.64); the multinomial logit estimate for someone who scored 1SD below the mean of sympathy and 1SD above the mean of inhibitory control was -3.50 (OR=0.03); the multinomial logit estimate for someone who scored 1SD above the mean of sympathy and 1SD below the mean of inhibitory control was -4.78 (OR=0.01); and the multinomial logit estimate for someone who scored 1SD above the mean of sympathy and inhibitory control was -2.36 (OR=0.09).



Interaction effect of age on the effect of moral emotions predicting the participant role of defender versus outsider

The multinomial logit estimate for someone who scored 1SD below the mean of age and moral emotions was -1.90 ((-2.32+(-.64*-1)+(1.05*-1)+(.83*-1*-1)), corresponding OR=0.15); the multinomial logit estimate for someone who scored 1SD below the mean of age and 1SD above the mean of moral emotions was -1.46 (OR=0.23); the multinomial logit estimate for someone who scored 1SD above the mean of age and 1SD below the mean of moral emotions was -4.84 (OR=0.01); and the multinomial logit estimate for someone who scored 1SD above the mean of age and moral emotions was -1.08 (OR=0.34).

Appendix F

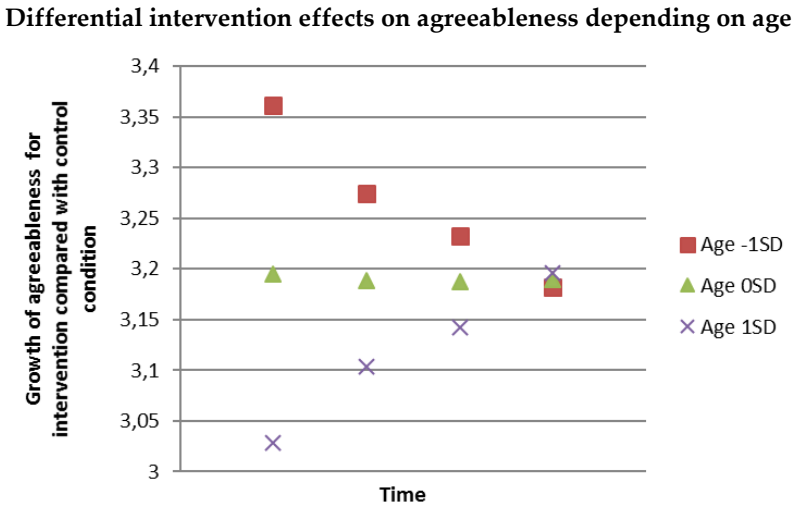


Figure 1. Interaction effect of age on growth of agreeableness for the intervention condition in comparison with the control condition over time

The estimate for the difference between agreeableness in the intervention condition and the control condition for someone at the first occasion (=1.46 SD below the mean of time) who scored 1SD below the mean of age was 3.36 (=3.452 + (0.081*-1.46) + (-0.004*-1.46*-1.46) + (-0.104*-1) + (-0.264*1) + (-0.083*-1.46*1) + (0.006*-1.46*-1.46*1) + (-0.01*-1.46*-1) + (0.001*-1.46*-1.46*-1) + (0.037*1*- 1) + (0.071*1*-1.46*-1) + (-0.004*1*-1.46*-1.46*-1). The estimate for someone who scored 1SD above the mean of age was 3.03 (=3.452 + (0.081*-1.46) + (-0.004*-1.46*-1.46) + (-0.104*1) + (-0.264*1) + (-0.083*-1.46*1) + (0.006*-1.46*-1.46*1) + (-0.01*-1.46*1) + (0.001*-1.46*-1.46*1) + (0.037*1*1) + (0.071*1*-1.46*1) + (-0.004*1*-1.46*-1.46*1). The estimate for the difference between agreeableness in the intervention condition and the control condition for someone at the last occasion (=1.38 SD above the mean of time) who scored 1SD below the mean of age was 3.18. The estimate for someone who scored 1SD above the mean of age was 3.20.

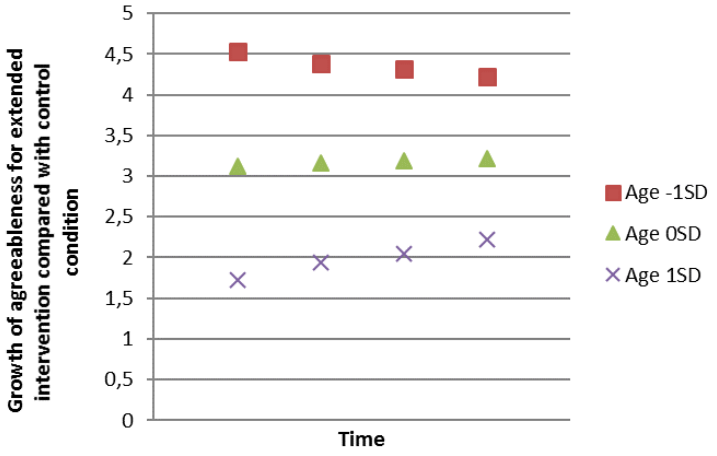


Figure 2. Interaction effect of age on growth of agreeableness for the extended intervention condition in comparison with the control condition over time

The estimate for the difference between agreeableness in the extended intervention condition and the control condition for someone at the first occasion (=1.46 SD below the mean of time) who scored 1SD below the mean of age was 4.53 (= 3.452 + (0.081*-1.46) + (-0.004*-1.46*-1.46) + (-0.104*-1) + (-0.28*1) + (-0.047*-1.46*1) + (0.005*-1.46*-1.46*1) + (-0.01*-1.46*-1) + (-0.001*-1.46*-1.46*-1) + (-1.08*1*-1) + (0.151*1*-1.46*-1) + (-0.007*1*-1.46*-1.46*-1). The estimate for someone who scored 1SD above the mean of age was 1.72 (= 3.452 + (0.081*-1.46) + (-0.004*-1.46*-1.46) + (-0.104*1) + (-0.28*1) + (-0.047*-1.46*1) + (0.005*-1.46*-1.46*1) + (-0.01*-1.46*1) + (-0.001*-1.46*-1.46*1) + (-1.08*1*1) + (0.151*1*-1.46*1) + (-0.007*1*-1.46*-1.46*1). The estimate for the difference between agreeableness in the extended intervention condition and the control condition for someone at the last occasion (=1.38 SD above the mean of time) who scored 1SD below the mean of age was 4.23. The estimate for someone who scored 1SD above the mean of age was 2.22.

Differential intervention effects on the other outcomes depending on gender

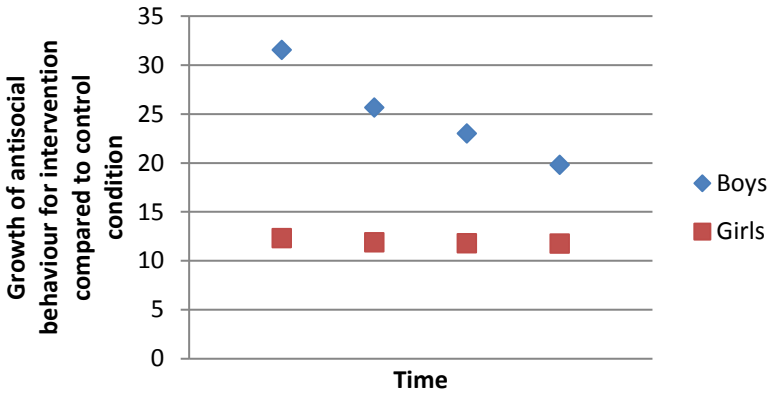


Figure 3. Interaction effect of gender on growth of antisocial behaviour for the intervention condition in comparison with the control condition over time

The estimate for the difference between antisocial behaviour in the intervention condition and the control condition for a boy at the first occasion (=1.46 SD below the mean of time) was 31.55 (20.032 + (-0.54*-1.46) + (-0.049*-1.46*-1.46) + (0.003*-1.46*-1.46*-1.46) + (-13.062*0) + (4.431*1) + (-3.516*-1.46*1) + (0.574*-1.46*-1.46*1) + (-0.021*-1.46*-1.46*-1.46*1) + (3.241*-1.46*0) + (-0.432*-1.46*-1.46*0) + (0.015*-1.46*-1.46*-1.46*0) + (0.459*1*0) + (0.642*1*-1.46*0) + (0*1*-1.46*-1.46*0) + (0*1*-1.46*-1.46*-1.46*0). The estimate for a girl was 12.32. The estimate for the difference between antisocial behaviour in the intervention condition and the control condition for a boy at the last occasion (=1.38 SD above the mean of time) was 19.83. The estimate for a girl was 11.79.

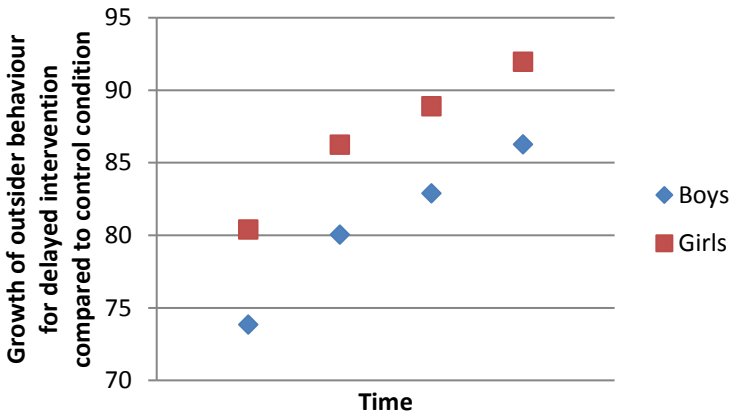


Figure 4. Interaction effect of gender on growth of outsider behaviour for the delayed intervention condition in comparison with the control condition over time

The estimate for the difference between outsider behaviour in the delayed intervention condition and the control condition for a boy at the first occasion (=1.46 SD below the mean of time) was 73.84 (89.368 + (-1.292*-1.46) + (0.285*-1.46*-1.46) + (-0.013*-1.46*-1.46*-1.46) + (1.553*0) + (-8.02*1) + (5.592*-1.46*1) + (-0.836*-1.46*-1.46*1) + (0.032*-1.46*-1.46*-1.46*1) + (0.125*-1.46*0) + (0*-1.46*-1.46*0) + (0*-1.46*-1.46*-1.46*0) + (4.562*1*0) + (-0.427*1*-1.46*0) + (0*1*-1.46*-1.46*0) + (0*1*-1.46*-1.46*-1.46*0). The estimate for a girl was 80.40. The estimate for the difference between outsider behaviour in the delayed intervention condition and the control condition for a boy at the last occasion (=1.38 SD above the mean of time) was 86.27. The estimate for a girl was 91.96.

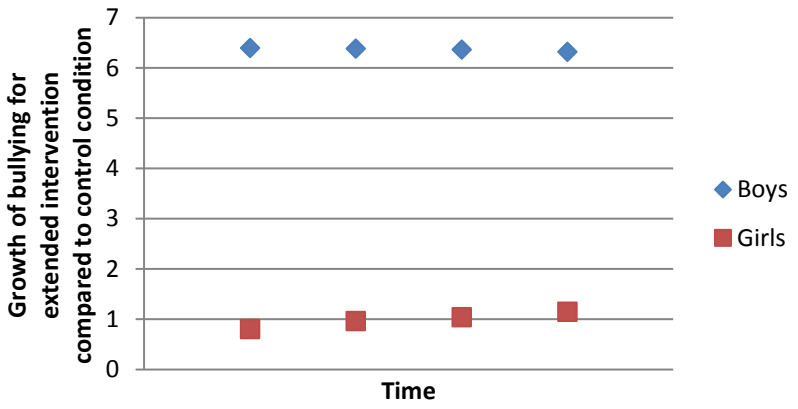


Figure 5. Interaction effect of gender on growth of bullying for the extended intervention condition in comparison with the control condition over time

The estimate for the difference between bullying in the extended intervention condition and the control condition for a boy at the first occasion (=1.46 SD below the mean of time) was 6.41 (2.588+ (-0.006*-1.46) + (-0.003*-1.46*-1.46) + (0*-1.46*-1.46*-1.46) + (-1.823*0) + (3.790*1) + (-0.022*-1.46*1) + (-0.001*-1.46*-1.46*1) + (0*-1.46*-1.46*-1.46*1) + (-0.046*-1.46*0) + (0*-1.46*-1.46*0) + (0*-1.46*-1.46*-1.46*0) + (-3.555*1*0) + (0.195*1*-1.46*0) + (0*1*-1.46*-1.46*0) + (0*1*-1.46*-1.46*-1.46*0). The estimate for a girl was 0.80. The estimate for the difference between bullying in the extended intervention condition and the control condition for a boy at the last occasion (=1.38 SD above the mean of time) was 6.32. The estimate for a girl was 1.15.

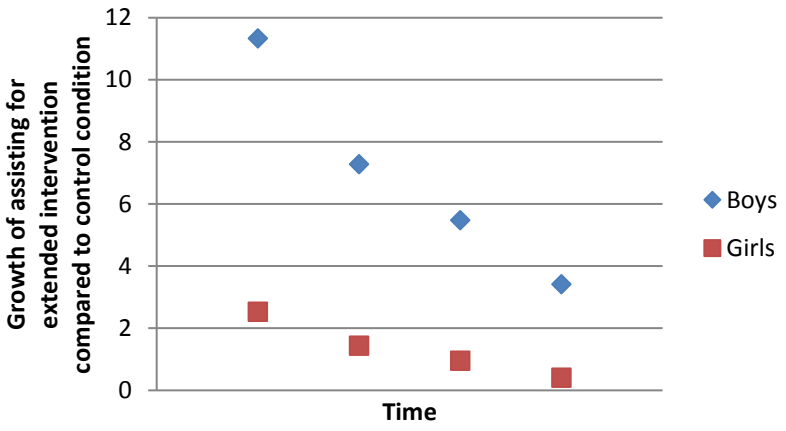


Figure 6. Interaction effect of gender on growth of assisting for the extended intervention condition in comparison with the control condition over time

The estimate for the difference between assisting in the extended intervention condition and the control condition for a boy at the first occasion (=1.46 SD below the mean of time) was 11.34 (2.047+ (-0.060*-1.46) + (-0.001*-1.46*-1.46) + (0.0001*-1.46*-1.46*-1.46) + (-1.537*0) + (4.401*1) + (-2.668*-1.46*1) + (0.406*-1.46*-1.46*1) + (-0.016*-1.46*-1.46*-1.46*1) + (-0.045*-1.46*0) + (0.014*-1.46*-1.46*0) + (-0.001*-1.46*-1.46*-1.46*0) + (-3.696*1*0) + (2.044*1*-1.46*0) + (-0.309*1*-1.46*-1.46*0) + (0.012*1*-1.46*-1.46*-1.46*0). The estimate for a girl was 2.53. The estimate for the difference between assisting in the extended intervention condition and the control condition for a boy at the last occasion (=1.38 SD above the mean of time) was 3.42. The estimate for a girl was 0.41.

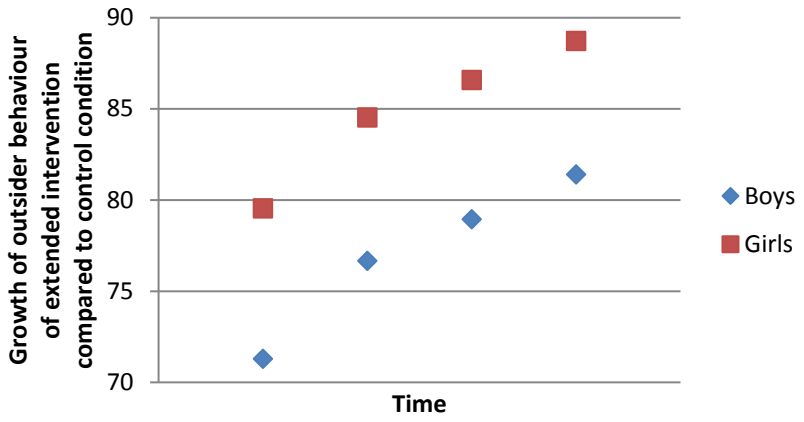


Figure 7. Interaction effect of gender on growth of outsider behaviour for the extended intervention condition in comparison with the control condition over time

The estimate for the difference between outsider behaviour in the extended intervention condition and the control condition for a boy at the first occasion (=1.46 SD below the mean of time) was 71.29 (89.368 + (-1.292*-1.46) + (0.285*-1.46*-1.46) + (-0.013*-1.46*-1.46*-1.46) + (1.553*0) + (-11.648*1) + (4.761*-1.46*1) + (-0.893*-1.46*-1.46*1) + (0.039*-1.46*-1.46*-1.46*1) + (0.125*-1.46*0) + (0*-1.46*-1.46*0) + (0*-1.46*-1.46*-1.46*0) + (6.218*1*0) + (-0.455*1*-1.46*0) + (0*1*-1.46*-1.46*0) + (0*1*-1.46*-1.46*-1.46*0). The estimate for a girl was 79.54. The estimate for the difference between outsider behaviour in the extended intervention condition and the control condition for a boy at the last occasion (=1.38 SD above the mean of time) was 81.41. The estimate for a girl was 88.73.

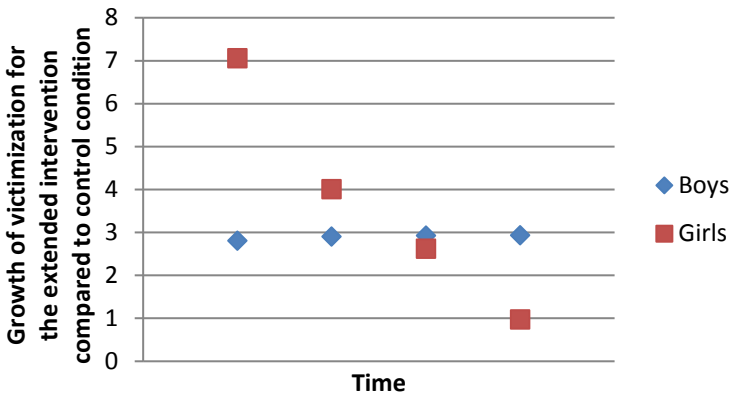


Figure 8. Interaction effect of gender on growth of victimization for the extended intervention condition in comparison with the control condition over time

The estimate for the difference between victimization in the extended intervention condition and the control condition for a boy at the first occasion (=1.46 SD below the mean of time) was 2.81 (2.025+ (-0.183*-1.46) + (-0.014*-1.46*-1.46) + (0.001*-1.46*-1.46*-1.46) + (0.137*0) + (0.891*1) + (0.224*-1.46*1) + (-0.006*-1.46*-1.46*1) + (-0.0001*-1.46*-1.46*-1.46*1) + (-0.053*-1.46*0) + (-0.03*-1.46*-1.46*0) + (0.002*-1.46*-1.46*-1.46*0) + (0.284*1*0) + (-2.094*1*-1.46*0) + (0.326*1*-1.46*-1.46*0) + (-0.012*1*-1.46*-1.46*-1.46*0). The estimate for a girl was 7.06. The estimate for the difference between victimization in the extended intervention condition and the control condition for a boy at the last occasion (=1.38 SD above the mean of time) was 2.94. The estimate for a girl was 0.98.

Differential intervention effects on the other outcomes depending on age

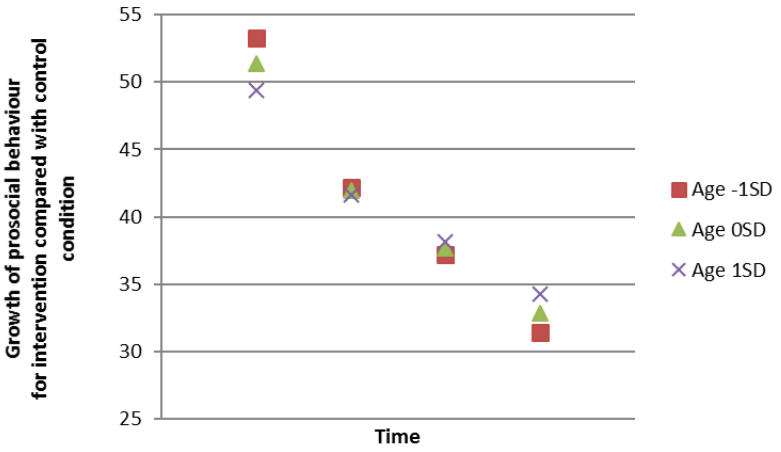


Figure 9. Interaction effect of age on growth of prosocial behaviour for the intervention condition in comparison with the control condition over time

The estimate for the difference between prosocial behaviour in the intervention condition and the control condition for someone at the first occasion (=1.46 SD below the mean of time) who scored 1SD below the mean of age was 53.27 (= 41.442 + (-2.025*-1.46) + (0.264*-1.46*-1.46) + (-0.006*-1.46*-1.46*-1.46) (-1.693*-1) + (-1.478*1) + (-4.373*-1.46*1) + (0.652*-1.46*-1.46*1) + (-0.026*-1.46*-1.46*-1.46*1) + (3.630*-1.46*-1) + (-0.551*-1.46*-1.46*-1) + (0.019*-1.46*-1.46*-1.46*-1) + (1.732*1*-1) + (-2.460*1*-1.46*-1) + (0.432*1*-1.46*-1.46*-1)+ (0.432*1*-1.46*-1.46*-1.46*-1)). The estimate for someone who scored 1SD above the mean of age was 49.41. The estimate for the difference between prosocial behaviour in the intervention condition and the control condition for someone at the last occasion (=1.38 SD above the mean of time) who scored 1SD below the mean of age was 31.38. The estimate for someone who scored 1SD above the mean of age was 34.25.

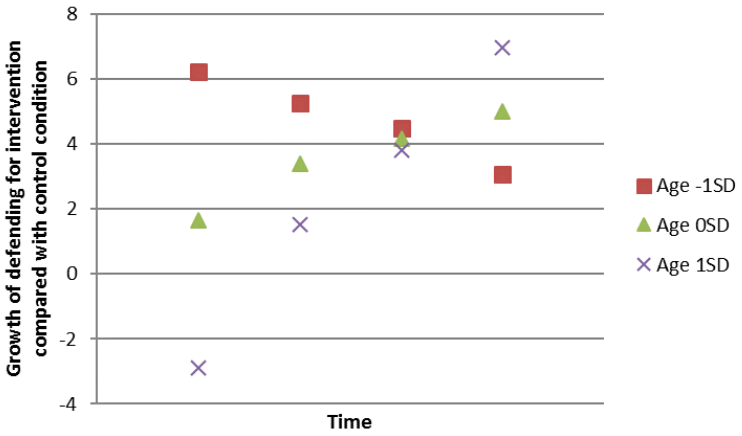


Figure 10. Interaction effect of age on growth of defending for the intervention condition in comparison with the control condition over time

The estimate for the difference between defending in the intervention condition and the control condition for someone at the first occasion (=1.46 SD below the mean of time) who scored 1SD below the mean of age was 6.22 (3.933 + (1.301*-1.46) + (-0.214*-1.46*-1.46) + (0.009*-1.46*-1.46*-1.46) + (-0.528*-1) + (-0.185*1) + (-0.145*-1.46*1) + (0.037*-1.46*-1.46*1) + (-0.002*-1.46*-1.46*-1.46*1) + (1.157*-1.46*-1) + (-0.166*-1.46*-1.46*-1) + (0.006*-1.46*-1.46*-1.46*-1) + (-0.678*1*-1) + (1.141*1*-1.46*-1) + (0.165*1*-1.46*-1.46*-1)+ (-0.006*1*-1.46*-1.46*-1.46*-1)). The estimate for someone who scored 1SD above the mean of age was -2.90. The estimate for the difference between defending in the intervention condition and the control condition for someone at the last occasion (=1.38 SD above the mean of time) who scored 1SD below the mean of age was 3.07. The estimate for someone who scored 1SD above the mean of age was 6.97.

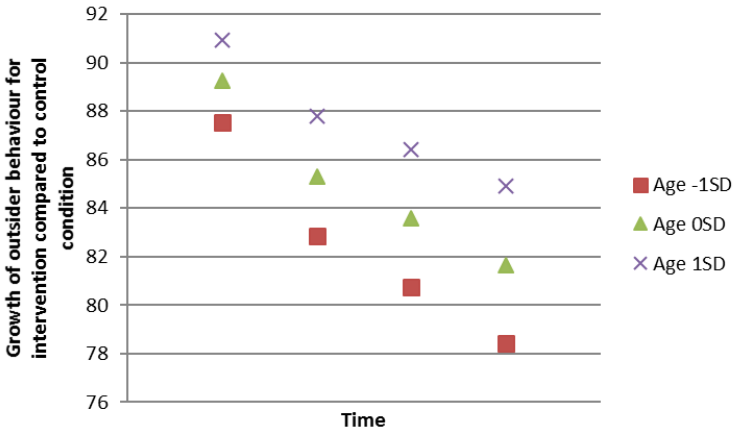


Figure 11. Interaction effect of age on growth of outsider behaviour for the intervention condition in comparison with the control condition over time

The estimate for the difference between outsider behaviour in the intervention condition and the control condition for someone at the first occasion (=1.46 SD below the mean of time) who scored 1SD below the mean of age was 87.55 (89.368 + (-1.292*-1.46) + (0.285*-1.46*-1.46) + (-0.013*-1.46*-1.46*-1.46) + (1.334*-1) + (-4.851*1) + (-1.314*-1.46*1) + (0.125*-1.46*-1.46*1) + (-0.003*-1.46*-1.46*-1.46*1) + (-0.810*-1.46*-1) + (0.117*-1.46*-1.46*-1) + (-0.004*-1.46*-1.46*-1.46*-1) + (1.304*1*-1) + (1.354*1*-1.46*-1) + (-0.185*1*-1.46*-1.46*-1)+ (0.006*1*-1.46*-1.46*-1.46*-1)). The estimate for someone who scored 1SD above the mean of age was 90.93. The estimate for the difference between outsider behaviour in the intervention condition and the control condition for someone at the last occasion (=1.38 SD above the mean of time) who scored 1SD below the mean of age was 78.40. The estimate for someone who scored 1SD above the mean of age was 84.93.

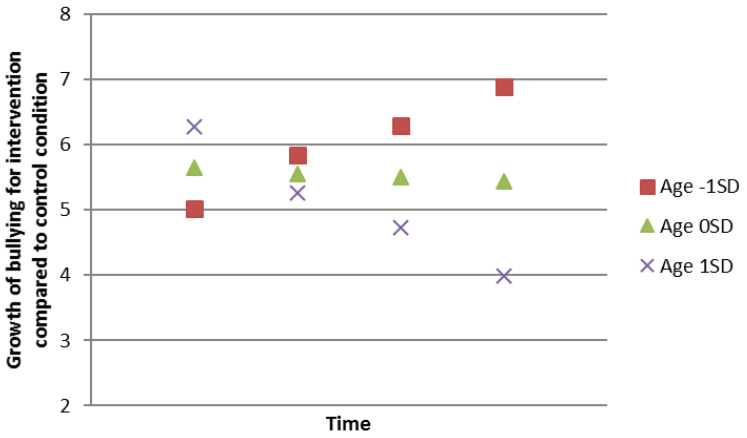


Figure 12. Interaction effect of age on growth of bullying for the delayed intervention condition in comparison with the control condition over time

The estimate for the difference between bullying in the intervention condition and the control condition for someone at the first occasion (=1.46 SD below the mean of time) who scored 1SD below the mean of age was 5.02 (2.588+ (-0.006*-1.46) + (0.003*-1.46*-1.46) + (0*-1.46*-1.46*-1.46) + (-0.238*-1) + (2.948*1) + (-0.067*-1.46*1) + (-0.001*-1.46*-1.46*1) + (0*-1.46*-1.46*-1.46*1) + (-0.012*-1.46*-1) + (0.001*-1.46*-1.46*-1) + (0*-1.46*-1.46*-1.46*-1) + (-0.257*1*-1) + (-0.717*1*-1.46*-1) + (0.032*1*-1.46*-1.46*-1) + (0*1*-1.46*-1.46*-1.46*-1). The estimate for someone who scored 1SD above the mean of age was 6.27. The estimate for the difference between bullying in the intervention condition and the control condition for someone at the last occasion (=1.38 SD above the mean of time) who scored 1SD below the mean of age was 6.89. The estimate for someone who scored 1SD above the mean of age was 3.99.

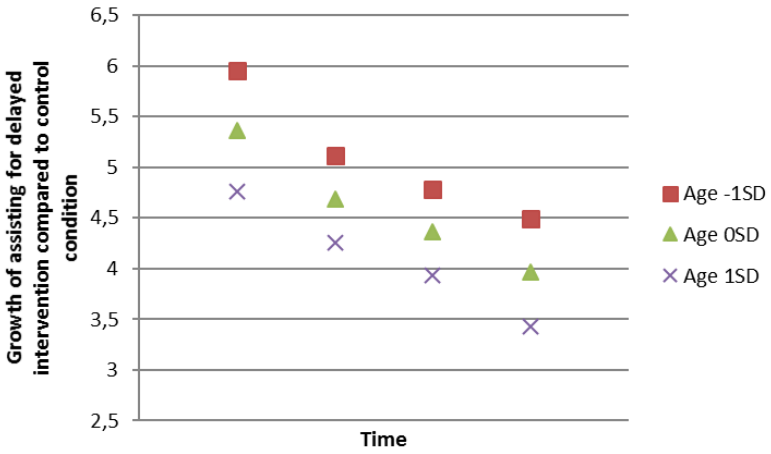


Figure 13. Interaction effect of age on growth of assisting for the delayed intervention condition in comparison with the control condition over time

The estimate for the difference between assisting in the delayed intervention condition and the control condition for someone at the first occasion (=1.46 SD below the mean of time) who scored 1SD below the mean of age was 5.95 (2.047 + (-0.060*-1.46) + (-0.001*-1.46*-1.46) + (0.0001*-1.46*-1.46*-1.46) + (-0.323*-1) + (2.489*1) + (-0.425*-1.46*1) + (0.053*-1.46*-1.46*1) + (-0.002*-1.46*-1.46*-1.46*1) + (-0.093*-1.46*-1) + (0.013*-1.46*-1.46*-1) + (-0.001*-1.46*-1.46*-1.46*-1) + (-0.096*1*-1) + (0.103*1*-1.46*-1) + (-0.084*1*-1.46*-1.46*-1)+ (0.004*1*-1.46*-1.46*-1.46*-1)). The estimate for someone who scored 1SD above the mean of age was 4.77. The estimate for the difference between assisting in the delayed intervention condition and the control condition for someone at the last occasion (=1.38 SD above the mean of time) who scored 1SD below the mean of age was 4.49. The estimate for someone who scored 1SD above the mean of age was 3.43.

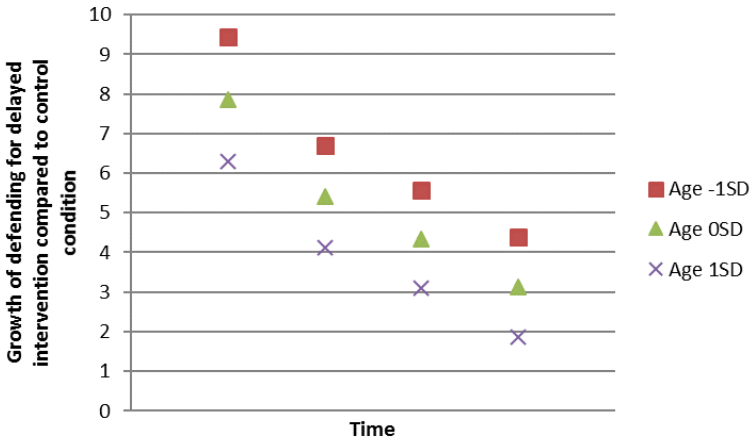


Figure 14. Interaction effect of age on growth of defending for the delayed intervention condition in comparison with the control condition over time

The estimate for the difference between defending in the delayed intervention condition and the control condition for someone at the first occasion (=1.46 SD below the mean of time) who scored 1SD below the mean of age was 9.44 (3.933 + (1.301*-1.46) + (-0.214*-1.46*-1.46) + (0.009*-1.46*-1.46*-1.46) + (-0.528*-1) + (0.976*1) + (-2.933*-1.46*1) + (0.472*-1.46*-1.46*1) + (-0.018*-1.46*-1.46*-1.46*1) + (1.157*-1.46*-1) + (-0.166*-1.46*-1.46*-1) + (0.006*-1.46*-1.46*-1.46*-1) + (-0.730*1*-1) + (-1.066*1*-1.46*-1) + (0.089*1*-1.46*-1.46*-1)+ (0.002*1*-1.46*-1.46*-1.46*-1)). The estimate for someone who scored 1SD above the mean of age was 6.29. The estimate for the difference between defending in the delayed intervention condition and the control condition for someone at the last occasion (=1.38 SD above the mean of time) who scored 1SD below the mean of age was 4.39. The estimate for someone who scored 1SD above the mean of age was 1.87.

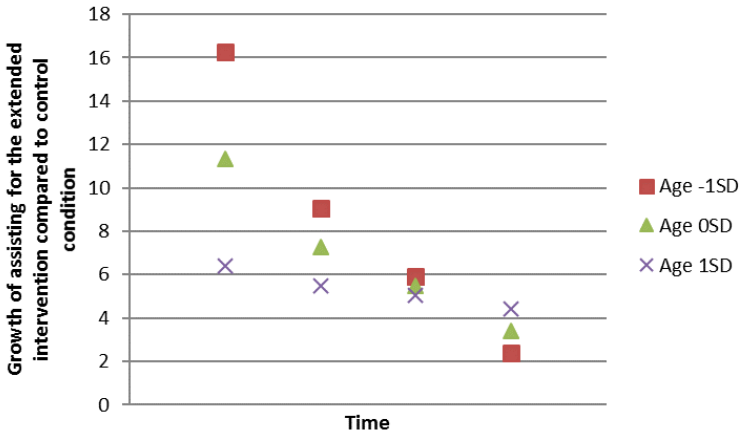


Figure 15. Interaction effect of age on growth of assisting for the extended intervention condition in comparison with the control condition over time

The estimate for the difference between assisting in the extended intervention condition and the control condition for someone at the first occasion (=1.46 SD below the mean of time) who scored 1SD below the mean of age was 16.26 (2.047 + (-0.060*-1.46) + (-0.001*-1.46*-1.46) + (0.0001*-1.46*-1.46*-1.46) + (-0.323*-1) + (4.401*1) + (-2.668*-1.46*1) + (0.406*-1.46*-1.46*1) + (-0.016*-1.46*-1.46*-1.46*1) + (-0.093*-1.46*-1) + (0.013*-1.46*-1.46*-1) + (-0.001*-1.46*-1.46*-1.46*-1) + (-0.847*1*-1) + (2.131*1*-1.46*-1) + (-0.357*1*-1.46*-1.46*-1)+ (0.014*1*-1.46*-1.46*-1.46*-1)). The estimate for someone who scored 1SD above the mean of age was 6.42. The estimate for the difference between assisting in the extended intervention condition and the control condition for someone at the last occasion (=1.38 SD above the mean of time) who scored 1SD below the mean of age was 2.41. The estimate for someone who scored 1SD above the mean of age was 4.43.

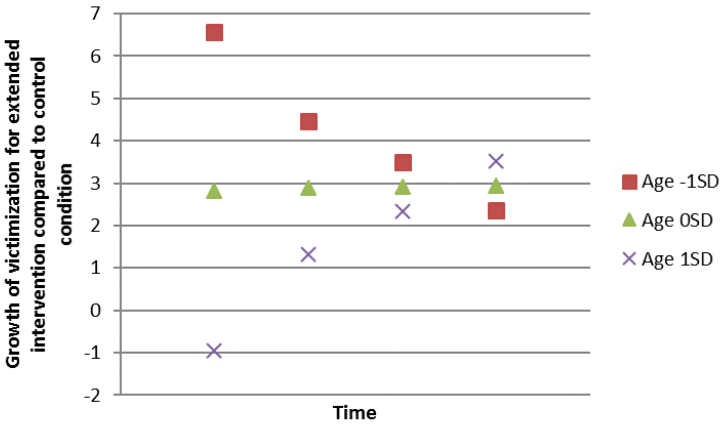


Figure 16. Interaction effect of age on growth of victimization for the extended intervention condition in comparison with the control condition over time

The estimate for the difference between victimization in the extended intervention condition and the control condition for someone at the first occasion (=1.46 SD below the mean of time) who scored 1SD below the mean of age was 6.58 (2.025 + (-0.183*-1.46) + (-0.014*-1.46*-1.46) + (0.001*-1.46*-1.46*-1.46) + (-0.241*-1) + (0.891*1) + (0.224*-1.46*1) + (-0.006*-1.46*-1.46*1) + (0.0001*-1.46*-1.46*-1.46*1) + (-0.204*-1.46*-1) + (0.029*-1.46*-1.46*-1) + (-0.001*-1.46*-1.46*-1.46*-1) + (-0.876*1*-1) + (1.709*1*-1.46*-1) + (-0.232*1*-1.46*-1.46*-1)+ (0.008*1*-1.46*-1.46*-1.46*-1)). The estimate for someone who scored 1SD above the mean of age was -0.96. The estimate for the difference between victimization in the extended intervention condition and the control condition for someone at the last occasion (=1.38 SD above the mean of time) who scored 1SD below the mean of age was 2.35. The estimate for someone who scored 1SD above the mean of age was 3.52.

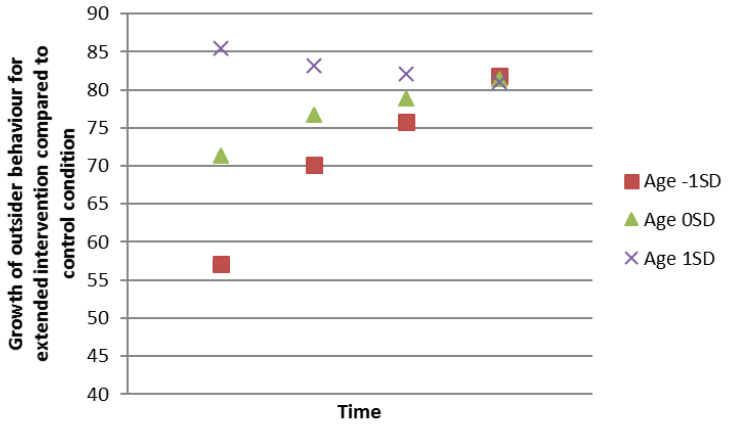


Figure 17. Interaction effect of age on growth of outsider behaviour for the extended intervention condition in comparison with the control condition over time

The estimate for the difference between outsider behaviour in the extended intervention condition and the control condition for someone at the first occasion (=1.46 SD below the mean of time) who scored 1SD below the mean of age was 57.08 (89.368 + (-1.292*-1.46) + (0.285*-1.46*-1.46) + (-0.013*-1.46*-1.46*-1.46) + (1.334*-1) + (-11.648*1) + (4.761*-1.46*1) + (-0.893*-1.46*-1.46*1) + (0.039*-1.46*-1.46*-1.46*1) + (-0.810*-1.46*-1) + (0.117*-1.46*-1.46*-1) + (-0.004*-1.46*-1.46*-1.46*-1) + (3.594*1*-1) + (-4.211*1*-1.46*-1) + (0.754*1*-1.46*-1.46*-1)+ (-0.031*1*-1.46*-1.46*-1.46*-1)). The estimate for someone who scored 1SD above the mean of age was 85.50. The estimate for the difference between outsider behaviour in the extended intervention condition and the control condition for someone at the last occasion (=1.38 SD above the mean of time) who scored 1SD below the mean of age was 81.83. The estimate for someone who scored 1SD above the mean of age was 80.99.

Appendix G

Two examples of lesson preparation forms

Lesson 2 preparation form – grade 3 to grade 6

Grade: grade 3 to 6
Lesson subject: Getting to know each other better
Date: …

Preparatory activities
The external teacher ensures that the materials are ready.

Initial situation
The children know where all the other children in their class live and what the house of their classmates looks like and can recognize them.

Learning goals

- The children can name their own characteristics and explain why they think these qualities fit them;
- The children can name characteristics they like and dislike in themselves;
- The children know and can name characteristics they recognize in their classmates;
- The children can explain why they think certain characteristics fit their classmates;
- The children know why it's important to know who they are and who someone else is
- The children know the benefits of getting to know others.

Instructional method
Instruction by teacher, play, independent work, group discussion

Materials:

- Music
- Firm A4-paper
- Pencils

Time	Activities children	Activities external teacher	Media
Feedback (5 min.)	The houses the children drew the previous week are	The external teachers leads to conversation.	

	briefly discussed. Children are asked to name the peer that drew the house the external teacher presents tot them (several al presented).		
Introduction (20 min.)	The previous lesson we dwelled upon the house and family of classmates. Today we are handling the topic of their own characteristics. Game 'How well do you know your classmates": In the game the children are spread out over the classroom. When the music starts, the children start walking through the classroom. When the music stops, children try to find a mate as soon as possible (within 3 seconds). They look closely at one another for 10 sec.. Then both children turn around. The teacher then asks one of the children questions, such as 'Does Jan have blue pants?'. After asking these questions, the music goes on. Every time the music stops, new questions are being asked. The game is discussed. Was it hard/ easy? Why?	The external teachers leads the game and asks questions.	Digiboard

	We have now talked about visible characteristics. Are there characteristics that are not visible? This is discussed briefly in the form of a spider web on the whiteboard. Both 'good' and 'less good' qualities are touched upon. Children are asked to think about their own characteristics. Children can also indicate why it is important to know the characteristics of each other. You always have an image of yourself. Sometimes other children have the same image of you, but sometimes this is also different. We are now going to play a game where the children think about which characteristics they find best suited for their peers. The children walk through the classroom and the external teacher calls statements with characteristics in them, e.g.: 'I admire you because you are a good drawer', 'I think you are very funny'. The children have to put their hand on the shoulder of the classmate they think the statement fits best.			

	We have now talked about all sorts of characteristics. First about what you think is a characteristic that suits you well and then characteristics that other children recognize in you.		
Core (25 min.)	These characteristics of the children they will draw in a drawing of their own hand. First, they must outline their own hand. In each finger, they put a characteristic they find fitting for themselves. The assignment that the children get is that they must note three good and two less good characteristics in their hand.	The external teacher indicates that the children will outline their own hand and that in each finger they have to write five characteristics that fit them most. The external teacher indicates that not only the word has to be written in the fingers, but that it must also be a beautiful drawing that we can add to the class book (which the children will do with the group teacher this week).	Digiboard
End (10 min.)	The children hand the external teacher the drawings of their hands. The children are called forward one at the time. The children read an random drawing and the other children guess who made it. Then, the external teacher asks the children to stand in class next to a classmate they don't know that well	The external teacher also indicates that the hands are all bundled in a classroom's friends book, and that the group's teacher also has an assignment this week to add to this. The external teacher explains that the form for the classroom's friends book is a robot with questions in it, like:	

	(according to the children themselves). These children stand opposite each other in the circle, so you get an inner and outer circle. Next, the inner circle slides one side aside. The children are told to fill in a form for a friends book. Not about themselves, but about the person they are currently facing.	hobbies, favorite food, etc. When the external teacher asks the children to choose a classmate they do not yet know, the children are not yet aware of what this exercise is about. When children know, they probably take this into account when choosing a classmate.	
THIS WEEK	The group teacher is going to let to children fill in the fill-in-sheet in the robot for a classmate and eventually add the robot to the classroom's friends book. The children should only complete one robot for a classmate. For example, I (Nilla) has to fill in the robot as if I'm Dorinde. In the top box, I then list 'My name is Dorinde'. Then I will answer all the questions, taking the perspective of Dorinde with every answer. When Nilla does not know something about Dorinde, she has to ask other classmates about Dorinde. Nilla is not allowed to ask Dorinde himself. The aim of this assignment is to stimulate pupils to think about the characteristics, hobbies and interests of other children. After filling in the robot, they will draw the same person (about whom they filled in the robot). Children should not only draw the face, but the whole body (look carefully at the colors of the clothes, shoes, hair, eyes, etc.). Again, the purpose of this assignment is that children think about peers and get to know each other better.		

Lesson 12 preparation form – grade 1 and grade 2

Grade: grade 1 and 2 Lesson subject: Trust Date: …
Preparatory activities The external teacher reserves a sports hall, games hall or an empty classroom and ensures that the materials are ready.

Initial situation

In the previous lessons, attention has been paid to getting to know each other better, rules, collaborating and emotions of themselves and of others.

Learning goals

- The children can indicate why (and for what specifically) they think it is important to trust one another;
- The children find out they need one another and that they have to trust one another to complete the game;
- The children can evaluate what went well and less well during the game and assignments;
- The children form a group in which they support one another and would give something up for one another.

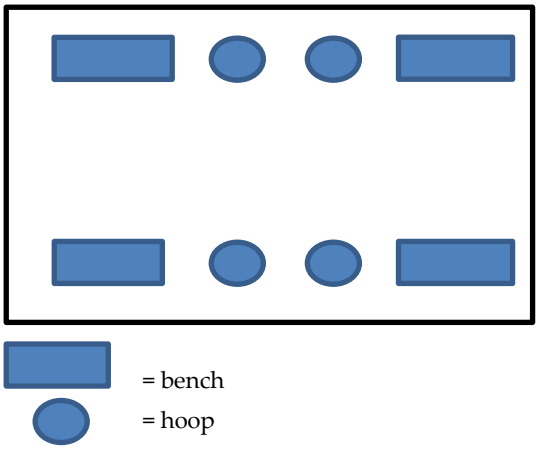
Instructional method

Group work and frontal teaching

Materials:

- Blindfolds
- 4 benches
- 4 hoops
- 2 mats

Time	**Activities children**	**Activities external teacher / Media**
Introduction (10 min.)	Exercise 1: The sports hall is empty. The children make couples on their own. The children probably choose someone that feel comfortable to them, and this is ok in the first exercise. One child of the pair now has to blindfold the other. They start any place in the sports hall. The child without blindfold guides the blindfolded child through the hall with his hands. The supervisor must guide the other so that they do not touch anyone or anything. The children are quiet during this exercise. After one minute the rolls are turned. This exercise can be played a second time, without touching during guidance; they guide each other by verbally giving directions.	

	Exercise 2: The external teacher creates two similar obstacle lanes (see illustration below). The external teacher makes couples of two (different pairs than in exercise 1). The couples are divided into two groups. Of each couple, one child gets a blindfold. The other has to guide this blindfolded peer over the lanes. Which group manages to be first with all the children on the other side? The external teacher ensures that all children have practiced both roles. **Rule**: the blindfolded child is not allowed to touch the floor. Setup exercise 2:  	
Core (40 min.)	The external teacher asks the children what they think of these exercises. Did they find it difficult to be blindfolded? Why? What did the children have to do during this exercise? The external teacher directs the children towards trust. A brief discussion is held about why and when trusting each other is important. The external teacher tells the children that during the previous exercises there were children who did not have a blindfold. Do the children dare to all be blindfolded? The game 'Who am I' is played: All children are blindfolded by the external teacher (the children can help). Then the children are placed anywhere in the hall (when they have their blindfold ready). When the	

	external teacher tells them, the children try to find another child and try to recognize each other. When the children think they know who they are, they stay together (children get 3-5 minutes for this assignment). Did the children guess right? How did the children find out? Did they like the exercise, or did the children find it scary / exciting, or ...? Then, the external teacher asks the children if they want to try something together (which is even more exciting than the blindfold assignments). The children probably answer 'yes', after which the external teacher puts the children in a circle (the equilibrium circle). All children form a circle, shoulder to shoulder. The children alternate with their face in and out of the circle. The children must hold on tight to their neighbors' wrists. When the external teacher calls 'start', all children lean forward, more and more, without taking a step forward. Does the equilibrium circle hold? On the sign 'stop' all children will be standing straight again. When the children want to, they can also lean backwards. The external teacher must first let the children stand straight again before the students release the wrists. *If the children do not dare to do the above exercise, the game "barriers" can be played. Children stand opposite of each other in a row with their arms stretched forward. One person walks through just before the children in the row raise their hands.*	
End (10 min.)	At the end of this lesson, a game is played, namely 'ant-tag'. This goes as follows: There are two children who tag. The taggers try to tag the other children. When a child is tagged, he/she lies down like an ant with arms and legs up and their back on the ground. The children who have not been tagged can take an arm or leg of an 'ant'. If there are four children, each holding an arm and leg of the other 'ant' child, they can carry the ant carefully to a mat. If the child is on the mat, he /she is free again. From the moment a child holds an arm or leg of an ant, he/she cannot be tagged. The game is over when all children are tagged.	

	Rules: - When holding a leg or arm of an ant you cannot be tagged - You can only lift an ant when you are with four children - You carry the ant carefully to the mat. Setup:   = mat	
THIS WEEK	The assignment in the context of trust for the group teacher is an Energizer that promotes team building in the classroom. The group teacher tells the children that they will be given several assignments and that they should perform it as soon as possible, e.g.: - Put two knees together - 3 left shoulders against each other - 5 feet - 3 heads - 5 noses (when the group teacher thinks this fits the group) In these exercises, it is important that the teacher takes into account the number of children in the group so that the exercises can be completed. The group teacher may also choose to participate.	

Samenvatting
(Dutch Summary)

1. Achtergrond van en aanleiding voor het onderzoek

Antisociaal gedrag is een groot probleem in het basisonderwijs. Antisociaal gedrag op school heeft niet alleen negatieve gevolgen voor leraren en betrokken leerlingen, maar ook voor de samenleving, zowel op financieel als sociaal gebied (Soepboer, Veenstra & Verhulst, 2006). Er is dan ook een toenemend besef dat het onderwijs een bijdrage zou moeten leveren aan het ontmoedigen van antisociaal gedrag oftewel gedrag dat een ander schaadt (Brown, Corrigan & Higgins-D'Alessandro, 2012; Fink & Slade, 2016; Rupp & Veugelers, 2003), en aan het bevorderen van prosociale waarden en gedrag, oftewel vrijwillig gedrag dat ten goede komt aan iemand anders (Eisenberg, Spinrad & Knafo, 2015). In de onderwijspraktijk wordt dit besef weerspiegeld door tal van interventieprogramma's gericht op pesten, prosociaal handelen en moreel gedrag (Reiman & Dotger, 2008; Smith, Ananiadou & Cowie, 2003; Smith, Cousins & Stewart, 2005). Onderzoek toont echter aan dat de effectiviteit van deze interventieprogramma's wisselt (Wienke, Anthonijsz, Abrahamse, Daamen & Nieuwboer, 2014; Willems, Denessen, Hermans & Vermeer, 2012). Daarnaast is er weinig bekend over hoe en wanneer de effectieve interventieprogramma's werken (Smith et al., 2004). Positieve effecten lijken namelijk verdeeld over verschillende typen interventies (Merrel et al., 2008). In het verlengde hiervan zijn Gravemeijer en Kirschner (2007) van mening dat onderzoek naar onderwijsinnovaties niet alleen gericht moet zijn op effectiviteit (wat werkt?), maar ook op het proces (hoe werkt het?).

Kortom, er is behoefte aan onderzoek naar de fundamentele processen die ten grondslag liggen aan anti- en prosociaal gedrag op de basisschool. Door meer inzicht te verkrijgen in deze fundamentele processen, kunnen educatieve interventieprogramma's gerichter worden ingezet voor het beïnvloeden van anti- en prosociaal gedrag. Een fundamenteel proces dat sterk samenhangt met anti- en prosociaal gedrag is moreel functioneren (Cuevas, 2011; Caravita, Gini & Pozzoli, 2012; Gasser & Keller, 2009; Hymel, Rocke-Henderson & Bonanno, 2005; Menesini & Camodeca, 2008; Perren & Gutzwiller-Helfenfinger, 2012). Moreel functioneren verwijst naar het psychologische proces dat ten grondslag ligt aan de reactie van iemand op een specifiek probleem, conflict, of dilemma dat een morele beslissing en moreel gedrag vereist (Tappan, 2006). Morele beslissingen en moreel gedrag worden bepaald door de mate waarin rekening wordt gehouden met het

vooropstellen van het belang of welzijn van de samenleving als geheel of van andere personen dan jezelf (Gewirth, 1984). Zowel anti- als prosociaal gedrag zijn primaire voorbeelden van moreel gedrag, omdat beide gedragingen direct invloed hebben op het welzijn van anderen (Fabes, Carlo, Kupanoff & Laible, 1999; Turiel, 1983; 1998). De afname van een coherent waardesysteem en de toenemende individualisering in de westerse samenleving maakt onderzoek naar moreel functioneren bovendien nog relevanter (Brown, Corrigan & Higgins-D'Alessandro, 2012; Fink & Slade, 2016; Rupp & Veugelers, 2003). Dit proefschrift richt zich dan ook op moreel functioneren als proces onderliggend aan anti- en prosociaal gedrag in het basisonderwijs.

Het Vier Componenten Model van Rest (1983; 1986) biedt momenteel het meest adequate kader om de onderliggende psychologische processen van moreel gedrag te onderzoeken. Daarom werd het Vier Componenten Model als theoretisch raamwerk voor het proefschrift gebruikt. De vier onderliggende psychologische processen van moreel gedrag, of componenten, zijn morele sensitiviteit, moreel redeneren, morele motivatie en moreel karakter. Al deze componenten zijn volgens Rest (1983; 1986) nodig om moreel gedrag te vertonen en verklaren dus mogelijk ook anti- en prosociaal gedrag. De eerste component is morele sensitiviteit en betreft het interpreteren van een situatie in morele termen, d.w.z. hoe het welzijn van een ander wordt beïnvloed door eventuele acties van jezelf. De tweede component is moreel redeneren en betreft het integreren van verschillende overwegingen om te bepalen wat de juiste handeling is. De derde component is morele motivatie en betreft het belang dat gehecht wordt aan morele waarden ten opzichte van andere waarden (gericht op jezelf). De vierde component is moreel karakter en betreft het vermogen om, ondanks dat er obstakels zijn, toch moreel te handelen (Rest, 1983; 1986; 1994).

1.1 Het onderzoek

Het onderzoek dat beschreven wordt in dit proefschrift bestond uit vier verschillende, maar aanvullende studies. Het onderzoek begon met het ontwikkelen en het analyseren van de benodigde meetinstrumenten. In hoofdstuk 2 ging de aandacht uit naar het meten van morele motivatie, omdat kleine variaties in het meten van morele motivatie een grote invloed kunnen hebben op de reacties van kinderen op de vragen uit het meetinstrument (Malti & Ongley, 2014; Nunner-Winkler, 2013). Morele motivatie werd gemeten door middel van geanticipeerde

emoties na hypothetische morele dilemma's. Dit specifieke meetinstrument werd onderzocht met het oog op de vraag of morele motivatie samenhangt met het morele domein en de leeftijd van de kinderen. Ook werd morele motivatie gerelateerd aan belangrijke criteriummaten, te weten anti- en prosociaal gedrag en sympathie. In de volgende twee studies werden morele sensitiviteit, moreel redeneren, morele motivatie en moreel karakter onderzocht in relatie tot de ontwikkeling van prosociaal gedrag en in relatie tot pesten, een veelvoorkomende vorm van antisociaal gedrag in het basisonderwijs. In hoofdstuk 3 werd gebruik gemaakt van de theorie van marginale afwijkingen van Caprara en collega's (1992) om de ontwikkeling van prosociaal gedrag te verklaren. We onderzochten in hoeverre initiële marginale afwijkingen in prosociaal gedrag samenhangen met de verdere ontwikkeling (in positieve en negatieve zin) van prosociaal gedrag en in welke mate dit samenhangt met de (accumulatie van) morele processen. In hoofdstuk 4 werden de morele componenten gerelateerd aan pestgedrag. Rekening houdend met het groepskarakter van pesten, werd gekeken naar de relatieve bijdrage van de morele componenten op zowel individueel als klasniveau in relatie tot vijf verschillende rollen die kinderen in kunnen nemen in een pestsituatie: pester, assistent, buitenstaander, verdediger en slachtoffer. In hoofdstuk 5 werden de effecten van een klasinterventie in het basisonderwijs bekeken gericht op de bevordering van mildheid, een aspect van moreel karakter. Specifiek werd de afname van pesten, assisteren bij het pesten, slachtofferschap, en antisociaal gedrag en de toename van verdedigen van het slachtoffer, het buitenstaan, en prosociaal gedrag onderzocht in de loop van twee schooljaren. Deze studie was vooral bedoeld om een bijdrage te leveren aan de behoefte aan meer kennis over wat effectief is in morele educatie en de preventie van pestgedrag en antisociaal gedrag.

2. Hoofdbevindingen

Hieronder worden per hoofdstuk de hoofdbevindingen uit het voorliggende proefschrift besproken.

Hoofdstuk 2. Het meten van geanticipeerde emoties na morele overtredingen

De eerste studie richtte zich op het meten van morele motivatie, oftewel geanticipeerde emoties die kinderen rapporteren na een hypothetische morele overtreding. We waren geïnteresseerd in de betrouwbaarheid en validiteit van het

instrument en in de relatie tussen morele motivatie en het morele domein en de leeftijd van de kinderen. Alhoewel deze manier van het meten van morele motivatie al een aantal decennia wordt gebruikt, is dit de eerste studie waarin de betrouwbaarheid en (aspecten van de) validiteit systematisch onderzocht wordt. Het instrument bestond uit zes scenario's verdeeld over drie morele domeinen: eerlijkheid (niet eerlijk winnen, niet aan belofte houden), het nalaten van prosociaal gedrag (niet delen, niet helpen) en pesten (verbaal pesten, fysiek pesten). Geanticipeerde emoties die de kinderen van zes tot dertien jaar rapporteerden na de morele overtreding in het hypothetische scenario werden gecodeerd als negatief (oftewel moreel) of positief. Aan de hand van het instrument kon morele motivatie betrouwbaar gemeten worden. Er bleek één factor ten grondslag te liggen aan de geanticipeerde emoties gemeten op basis van de zes scenario's. Daarnaast werden enkele indicaties gevonden voor de predictieve en concurrente validiteit van het meetinstrument. Opvallend was dat prosociaal gedrag alleen voor specifieke scenario's samenhing met geanticipeerde emoties, namelijk voor het nalaten van prosociaal gedrag (niet delen, niet helpen). Dit wijst op domein specifieke variatie in geanticipeerde emoties na hypothetische morele overtredingen en op het belang van de aansluiting tussen de situaties van de morele scenario's en het gedrag waar de interesse naar uit gaat. Sympathie hing samen met de geanticipeerde emoties in alle hypothetische scenario's. Er werd geen relatie gevonden tussen geanticipeerde emoties en antisociaal gedrag. In tegenstelling tot de verwachtingen varieerden de geanticipeerde emoties niet naar gelang de leeftijd van de kinderen.

Op basis van de resultaten van de eerste studie is besloten om in de tweede en derde studie de rol van de vier componenten te onderzoeken voor prosociaal respectievelijk antisociaal gedrag afzonderlijk. Hoofdstuk 3 richtte zich op de ontwikkeling van prosociaal gedrag in relatie tot (de accumulatie van) morele componenten. In hoofdstuk 4 werd vervolgens gekeken naar de rol van morele componenten bij een specifieke vorm van antisociaal gedrag, namelijk pestgedrag. Het gelijktijdig onderzoeken van het aandeel van alle morele componenten was een belangrijke aanvulling op eerder onderzoek naar het Vier Componenten Model van Rest (1983;1986).

Hoofdstuk 3. Marginale afwijkingen in de ontwikkeling van prosociaal gedrag en de rol van morele componenten

In hoofdstuk 3 werd onderzocht of (de accumulatie van) morele componenten een modererende rol speelt bij de invloed van initiële marginale afwijkingen in prosociaal gedrag (positief en negatief) op de ontwikkeling van prosociaal gedrag over de tijd. Initiële afwijkingen in prosociaal gedrag werden hierbij gedefinieerd als een standaardscore groter dan 0 en kleiner dan 1 *SD* boven de norm of kleiner dan 0 en groter dan 1 *SD* onder de norm van initieel prosociaal gedrag. Omdat eerder onderzoek vooral geïnteresseerd was in extreme groepen of continue dimensies van gedrag zijn marginale afwijkingen in gedrag nog weinig onderzocht (Caprara et al., 2007). Echter, de theorie van marginale afwijkingen stelt dat ook marginale afwijkingen in gedrag de potentie hebben om zich te ontwikkelen tot extreme afwijkingen in gedrag. Naar marginale afwijkingen in antisociaal gedrag is inmiddels (in beperkte mate) onderzoek gedaan, maar over marginale afwijkingen in prosociaal gedrag is nog weinig bekend. Deze studie is dan ook een van de eerste waarin marginale afwijkingen in prosociaal gedrag zijn onderzocht, waardoor nieuwe inzichten werden verkregen over de ontwikkeling van prosociaal gedrag van kinderen in de basisschoolleeftijd. We onderzochten of initiële marginale afwijkingen in prosociaal gedrag zich ontwikkelen tot grotere afwijkingen (in positieve en negatieve zin) in prosociaal gedrag. Uitgaande van het Vier Componenten Model, werd daarnaast gekeken naar de unieke en geaggregeerde effecten van individuele verschillen in moreel functioneren op deze ontwikkeling van prosociaal gedrag. Er werd dus gekeken naar de modererende rol van moreel functioneren in het voorspellen van de relatie tussen marginale afwijkingen in prosociaal gedrag en de ontwikkeling van prosociaal gedrag. De resultaten lieten zien dat marginale afwijkingen in prosociaal gedrag zich niet ontwikkelden tot grotere afwijkingen in prosociaal gedrag over de tijd, ook niet in combinatie met de (accumulatie van) morele componenten. In het algemeen ontwikkelden initieel marginaal prosociale en marginaal nonprosociale kinderen zich naar het gemiddelde van prosociaal gedrag. De ontwikkeling van marginale afwijkingen in prosociaal gedrag tot meer prosociaal gedrag vond alleen maar plaats bij oudere kinderen met een hoge morele motivatie. Afwijkingen van de norm van prosociaal gedrag bleek dus voornamelijk gecompenseerd te worden door gedrag in de tegengestelde richting, in plaats van te leiden tot een versterking in de afwijkingen in prosociaal gedrag. Dit resultaat betekent wellicht dat de theorie

van marginale afwijkingen voornamelijk van toepassing is op gedrag dat snel opgemerkt wordt, zoals agressief of pestgedrag. Verder bleek de ontwikkeling van prosociaal gedrag niet voorspeld of gemodereerd te worden door morele sensitiviteit, moreel redeneren, of moreel karakter. Het verband tussen de accumulatie van morele componenten en de ontwikkeling van prosociaal gedrag was sterker dan het verband tussen de afzonderlijke componenten en de ontwikkeling van prosociaal gedrag. Dit suggereert dat kinderen hoog moeten scoren op meerdere morele componenten in plaats van op één morele component om zich positief te ontwikkelen op het gebied van prosociaal gedrag. Dit komt overeen met het Vier Componenten Model dat ervan uit gaat dat alle vier morele componenten nodig zijn om moreel te handelen.

Hoofdstuk 4. De relatie tussen pestgedrag en morele componenten op individueel en klasniveau

In hoofdstuk 4 werd gekeken naar de rol die de morele componenten (morele sensitiviteit, moreel redeneren, morele motivatie en moreel karakter) spelen in pestgedrag. Pesten is een ernstige vorm van antisociaal gedrag waarin een dader herhaaldelijk schade berokkent bij een slachtoffer. De gevolgen voor het slachtoffer kunnen grote proporties aannemen. Om recht te doen aan het fenomeen van pesten als groepsproces, werden de rollen van pester, assistent, verdediger, buitenstaander en slachtoffer vergeleken op het gebied van morele sensitiviteit, moreel redeneren, morele motivatie en moreel karakter op zowel individueel als klasniveau. Aansluitend bij het Vier Componenten Model werden alle morele componenten gelijktijdig bekeken, een aanvulling op eerdere studies die vaak niet meer dan twee componenten onderzochten. Bovendien werd de invloed van de morele componenten uitgediept door ze ook op klasniveau mee te nemen in de analyse. Uit de resultaten bleek dat alle morele componenten gerelateerd waren aan de kans om een van de rollen in te nemen in een pestsituatie. Dit wijst erop dat een geïntegreerde analyse van de vier verschillende morele componenten noodzakelijk is voor een beter begrip van de onderliggende processen van pestgedrag. Ook ondersteunt het het belang van het maken van een onderscheid tussen de verschillende pestrollen van pester, assistent, verdediger, buitenstaander en slachtoffer. Pesters bleken laag te scoren op nauwgezetheid, een onderdeel van moreel karakter, pesters en assistenten bleken hoog te scoren op moreel redeneren, en oudere verdedigers bleken hoog te scoren op morele sensitiviteit en op morele

motivatie ten opzichte van buitenstaanders. Opvallend was dat mildheid, ook onderdeel van moreel karakter, het meeste (negatief) samenhing met zowel pesten als assisteren in vergelijking met verdedigen en buitenstaan. Het stimuleren van mildheid op de basisschool lijkt daarom een veelbelovende weg om pestgedrag tegen te gaan. Tot slot werd pestgedrag voorspeld door een aantal morele componenten op klasniveau. Dit suggereert dat interventies er goed aan doen zich ook te richten op de gehele klas.

Hoofdstuk 5. Effecten van een klasinterventie gericht op het stimuleren van mildheid op de basisschool

Aangezien de vorige studie aantoonde dat mildheid de grootste (negatieve) samenhang vertoonde met pestgedrag, en eerder onderzoek soortgelijke resultaten liet zien, redeneerden we dat mildheid een succes bevorderende factor kon zijn voor educatieve interventies gericht op het tegengaan van pestgedrag en antisociaal gedrag op de basisschool. Mildheid bleek van alle persoonlijkheidseigenschappen bovendien het meest beïnvloedbaar door de omgeving (Bergeman et al., 1993; Graziano & Eisenberg, 1997). Het doel van hoofdstuk 5 was daarom om te onderzoeken hoe een educatieve interventie mildheid zou kunnen stimuleren om zo pestgedrag en antisociaal gedrag te verminderen en prosociaal gedrag te bevorderen. Door het onderzoek te richten op mildheid als potentieel succes bevorderende factor van de interventie, werd informatie verkregen over het verklarend proces achter de werkzaamheid van een interventie (Leidt het stimuleren van mildheid tot een afname van antisociaal gedrag en een toename van prosociaal gedrag?) alsook indicaties voor de effectiviteit ervan (Leidt de interventie tot een afname van antisociaal gedrag en een toename van prosociaal gedrag?). Meer specifiek werden de effecten onderzocht van een klasinterventie gericht op het bevorderen van mildheid bij kinderen tussen de zes en dertien jaar op zowel pestgedrag als pro- en antisociaal gedrag. De effecten van de klasinterventie werden gedurende twee schooljaren getoetst door kinderen in een controleconditie te vergelijken met kinderen in drie verschillende interventiecondities: (a) een interventieconditie waarin de klasinterventie tijdens het eerste jaar geïmplementeerd werd, (b) een uitgestelde interventieconditie waarin de klasinterventie tijdens het tweede jaar geïmplementeerd werd, en (c) een uitgebreide interventieconditie waarin de klasinterventie tijdens het eerste en het tweede jaar geïmplementeerd werd. Een bemoedigend en interessant resultaat is

dat we erin slaagden mildheid te stimuleren in alle drie de interventiecondities. Zoals verwacht verminderde het pestgedrag en antisociaal gedrag ook in alle interventiecondities. In vergelijking met de controleconditie waren de kinderen in alle interventiecondities na twee jaar meer gedaald in antisociaal gedrag, pesten, assisteren en slachtofferschap, en meer gestegen in buitenstaan. Hierbij werden een opvallend aantal leeftijds- en geslachts-specifieke effecten gevonden. De genoemde positieve effecten van de interventie waren bovendien het meest zichtbaar in de uitgebreide interventieconditie. De rol van mildheid in het voorspellen van de positieve effecten van de interventie was echter klein. Daarom is niet zeker of de verandering in mildheid de oorzaak is van de andere positieve effecten van de klasinterventie. Wel zijn er indicaties dat het stimuleren van mildheid in het basisonderwijs de potentie heeft om negatieve ontwikkelingstrajecten om te buigen.

3. Opbrengsten van het onderzoek

De vier deelstudies tezamen hebben, ondanks de beperkingen die in het proefschrift zijn beschreven, een aantal belangrijke opbrengsten opgeleverd. In de eerste plaats bleek dat moreel functioneren een belangrijk fundamenteel proces is in de ontwikkeling van pro- en antisociaal gedrag op de basisschool. De belangrijkste bijdrage van dit onderzoek is dan ook de identificatie van potentiële succes bevorderende (morele) componenten voor het ontmoedigen van antisociaal en pestgedrag en het bevorderen van prosociaal gedrag bij kinderen tussen de zes en dertien jaar. Onze resultaten lieten verder zien dat niet alle morele componenten in gelijke mate samenhangen met anti- en prosociaal gedrag. De sterkste relaties werden gevonden tussen moreel karakter en anti- en prosociaal gedrag. Daarnaast bleek dat de accumulatie en combinatie van morele componenten ook bijdroeg aan pro- en antisociaal gedrag, bovenop op de bijdragen van de individuele morele componenten. Naast individuele processen bleken groepsprocessen een belangrijke rol te spelen in de relatie tussen moreel functioneren en pro- en antisociaal gedrag. Praktijkbeoefenaars worden dan ook geadviseerd rekening te houden met klassen- en groepsprocessen bij het ontmoedigen van antisociaal gedrag danwel het bevorderen van prosociaal gedrag. Echter, alle bovengenoemde processen hadden een andere bijdrage aan prosociaal gedrag dan aan antisociaal gedrag. Dit betekent dat het tegengaan van antisociaal gedrag in het onderwijs een andere focus vereist dan het bevorderen van prosociaal gedrag. Aanvullend onderzoek naar de bijdrage

van morele componenten op individueel en klasniveau is echter nodig om meer inzicht te krijgen in de rol van de afzonderlijke componenten en de combinatie daarvan voor het bevorderen van prosociaal gedrag en het ontmoedigen van antisociaal gedrag.

Een tweede opbrengt van ons onderzoek betreft het inzicht in het oorzakelijk verband tussen mildheid en anti- en prosociaal gedrag. Door de vierde studie te richten op mildheid als potentieel succes-bevorderende factor van een klasinterventie, werd informatie verkregen over het verklarend proces achter de werkzaamheid van een interventie (Leidt het stimuleren van mildheid tot een afname van antisociaal gedrag en een toename van prosociaal gedrag?) alsook indicaties voor de effectiviteit ervan (Leidt de interventie tot een afname van antisociaal gedrag en een toename van prosociaal gedrag?). Het gebruik van drie verschillende interventiecondities en een longitudinaal design was bovendien een uitzonderlijk sterk punt van ons onderzoek. Zoals verwacht bleek mildheid een veranderbaar construct met het potentieel om pesten en antisociaal gedrag op de basisschool te ontmoedigen, vooral wanneer de interventie langer werd uitgevoerd. Verder experimenteel onderzoek zou de invloed van andere morele componenten, zoals bijvoorbeeld morele sensitiviteit, kunnen isoleren om zo de (causale) effecten van andere morele componenten op de ontwikkeling van pro- en antisociaal gedrag te achterhalen. Het door ons ontwikkelde interventieprogramma kan echter als vertrekpunt worden genomen voor de verdere verbetering van educatieve interventieprogramma's gericht op het tegengaan van antisociaal gedrag en pestgedrag. Het interventieprogramma geeft namelijk inzicht in de veranderingen die nodig zijn om mildheid te stimuleren op een manier die past bij de ontwikkeling van de kinderen, om zo antisociaal gedrag al op jonge leeftijd terug te dringen. We hopen dat deze, en andere uitkomsten van het onderzoek gebruikt kunnen worden om de waardevolle praktijkbeoefening en het onderzoek naar de rol van moreel functioneren in pro- en antisociaal gedrag te bevorderen.

References

Ahadi, S. A. & Rothbart, M. K. (1994). *Temperament, development, and the Big Five.* In: C. F. Halverson, G. A. Kohnstamm & R. P. Martin (Eds.), The Developing Structure of Temperament and Personality From Infancy To Adulthood (pp. 189-209). New York: Lawrence Erlbaum Associates, Inc.

Allport, G.W. (1954). *The nature of prejudice.* Cambrigde: Addison-Wesley.

Arsenio, W. F. (1988). Children's conceptions of the situational affective consequences of sociomoral events. *Child Development*, 56(6), 1611-1622.

Arsenio, W. F. (2014). Moral emotion attributions and aggression. In M. Killen & J.G. Smetana (Eds.), *Handbook of Moral Development* (pp. 235-255). New York: Psychology Press.

Arsenio, W. F. & Fleiss, K. (1996). Typical and behaviourally disruptive children's understanding of the emotional consequences of socio-moral events. *British Journal of Developmental Psychology*, *14*(2), 173-186.

Arsenio, W. F. & Ford, M. E. (1985). The role of affective information in social-cognitive development: Children's differentiation of moral and conventional events. *Merrill-Palmer Quarterly*, *31*(1), 1-17.

Arsenio, W., Gold, J. & Adams, E. (2006). Children's conceptions and displays of moral emotions. In M. Killen & J. Smetana (Eds.), *Handbook of Moral Development* (pp. 581–610). Mahwah, NJ: Erlbaum.

Arsenio, W. F. & Lemerise, E. A. (2001). Varieties of childhood bullying: Values, emotion processes, and social competence. *Social Development, 10*(1), 59-73.

Arsenio, W. F. & Lemerise, E. A. (2004). Aggression and moral development: Integrating social information processing and moral domain models. *Child Development, 75*(4), 987-1002.

Arsenio, W. & Lover, A. (1995). Children's conceptions of sociomoral affect: Happy victimizers, mixed emotions, and other expectancies. In M. Killen & D. Hart (Eds.), *Morality in everyday life: Developmental perspectives* (pp. 87-128). Cambridge: Cambridge University Press.

Asendorpf, J. B. & Nunner-Winkler, G. (1992). Children's moral motive strength and temperamental inhibition reduce their immoral behavior in real moral conflicts. *Child Development, 63*(5), 1223-1235.

Ashton, M. C., Lee, K. & De Vries, R. E. (2014). The HEXACO Honesty-Humility, Agreeableness, and Emotionality factors: A review of research and theory. *Personality and Social Psychology Review*, *18*(2), 139-152.

Atlas, R. S. & Pepler, D. J. (1998). Observations of bullying in the classroom. *The Journal of Educational Research, 92*(2), 86-99.

Bandura, A. (1986) *Social foundations of thought and action: A social-cognitive view.* Englewood Cliffs, NJ: Prentice-Hall.

Baldry, A. C. & Farrington, D. P. (2007). Effectiveness of programs to prevent school bullying. *Victims and Offenders*, *2*(2), 183-204.

Barbaranelli, C., Caprara, G. V., Rabasca, A. & Pastorelli, C. (2003). A questionnaire for measuring the Big Five in late childhood. *Personality and Individual Differences*, *34*(4), 645-664.

Barchia, K. & Bussey, K. (2011). Individual and collective social cognitive influences on peer aggression: Exploring the contribution of aggression efficacy, moral disengagement, and collective efficacy. *Aggressive Behavior, 37*(2), 107-120.

Barden, R. C., Zelko, F. A., Duncan, S. W. & Masters, J. C. (1980). Children's consensual knowledge about the experiential determinants of emotion. *Journal of Personality and Social Psychology*, *39*(5), 968.

Batson, C. D., Kobrynowincz, D., Dinnerstein, J. L., Kampf, H.C. & Wilson, A. D. (1997). In a very different voice: Unmasking moral hypocrisy. *Journal of Personality and Social Psychology, 72*, 1335-1348.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C. & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323.

Beauchaine, T. P., Gatzke-Kopp, L., Neuhaus, E., Chipman, J., Reid, M. J. & Webster-Stratton, C. (2013). Sympathetic-and parasympathetic-linked cardiac function and prediction of externalizing behavior, emotion regulation, and prosocial behavior among preschoolers treated for ADHD. *Journal of Consulting and Clinical Psychology*, *81*(3), 481.

Bebeau, M. J. 1994. Influencing the moral dimension of dental practice. In J. Rest & D. Narvaez (Eds.) *Moral Development in the Professions* (pp. 121-146). Hillsdale, NJ:Lawrence Erlbaum.

Bebeau, M. J., Rest, J. R. & Narvaez, D. (1999). Beyond the promise: A perspective on research in moral education. *Educational Researcher, 28*(4), 18-26.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238 -246.

Bergeman, C. S., Chlpuer, H. M., Plomin, R., Pedersen, N. L., McClearn, G. E., Nesselroade, J. R., ... & McCrae, R. R. (1993). Genetic and environmental effects on openness to experience, agreeableness, and conscientiousness: An adoption/twin study. *Journal of Personality*, *61*(2), 159-179.

Berkowitz, M. W., Sherblom, S., Bier, M. & Battistich, V. (2006). Educating for positive youth development. *Handbook of Moral Development*, 683-701.

Björkqvist, K., Österman, K. & Kaukiainen, A. (2000). Social intelligence− empathy= aggression?. *Aggression and Violent Behavior, 5*(2), 191-200.

Blasi, A. (1980). Bridging moral cognition and moral action: a critical review of the literature. *Psychological Bulletin, 88*, 1–45.

Bollmer, J. M., Harris, M. J. & Milich, R. (2006). Reactions to bullying and peer victimization: Narratives, physiological arousal, and personality. *Journal of Research in Personality, 40*(5), 803-828.

Brown, P., Corrigan, M. W. & Higgins-D'Alessandro, A. (Eds.). (2012). *Handbook of Prosocial Education (Vol. 1).*Lanham: Rowman & Littlefield.

Bustamante, A. & Chaux, E. (2014). Reducing moral disengagement mechanisms: A comparison of two interventions. *Journal of Latino/Latin American Studies*, *6*(1), 52-54.

Camodeca, M. & Goossens, F. A. (2005). Aggression, social cognitions, anger and sadness in bullies and victims. *Journal of Child Psychology and Psychiatry, 46*(2), 186-197.

Caprara G.V. (1992). Marginal deviations, aggregated effects, disruption of continuity, and deviation amplifying mechanisms. In: Hettema P.J., Deary I.J. (Eds.), *Foundations of Personality* (pp. 227–244). Dordrecht: Kluwer Academic.

Caprara, G. V., Dodge, K. A., Pastorelli, C., Zelli, A. (2007). The Conduct Problems Prevention Research Group Effect of marginal deviations on behavioral development and social judgment. *European Psychologist, 11,* 79–89.

Caprara, G. V., Dodge, K. A., Pastorelli, C., Zelli, A. (2007). How marginal deviations sometimes grow into serious aggression. *Child Development Perspectives, 1*, 33–39.

Caprara, G. V., Kanacri, B. P. L., Gerbino, M., Zuffianò, A., Alessandri, G., Vecchio, G., ... & Bridglall, B. (2014). Positive effects of promoting prosocial behavior in early adolescence: Evidence from a school-based intervention. *International Journal of Behavioral Development*, *38*(4), 386-396.

Caprara G.V. & Pastorelli C. (1993). Quando Poco e' Tanto. *Giornale Italiano di Psicologia, 20*, 427–451.

Caprara, G. I. A. N. & Zimbardo, P. G. (1996). Aggregation and amplification of marginal deviations in the social construction of personality and maladjustment. *European Journal of Personality, 10*(2), 79-110.

Caravita, S., DiBlasio, P. & Salmivalli, C. (2010). Participation in bullying among early adolescents: Is ToM involved? *Journal of Early Adolescence, 30(1),* 138-170.

Caravita, S., Gini, G. & Pozzoli, T. (2012). Main and moderated effects of moral cognition and status on bullying and defending. *Aggressive Behavior, 38*(6), 456-468.

Carlo, G., Crockett, L. J., Wolff, J. M. & Beal, S. J. (2012). The role of emotional reactivity, self-regulation, and puberty in adolescents' prosocial behaviors. *Social Development*, *21*(4), 667-685.

Carlo, G., Hausmann, A., Christiansen, S. & Randall, B. A. (2003). Sociocognitive and behavioral correlates of a measure of prosocial tendencies for adolescents. *The Journal of Early Adolescence*, *23*(1), 107-134.

Carlo, G., Mestre, M. V., McGinley, M. M., Samper, P., Tur, A. & Sandman, D. (2012). The interplay of emotional instability, empathy, and coping on prosocial and aggressive behaviors. *Personality and Individual Differences, 53*(5), 675-680.

Carlo, G., Mestre, M. V., Samper, P., Tur, A. & Armenta, B. E. (2011). The longitudinal relations among dimensions of parenting styles, sympathy, prosocial moral reasoning, and prosocial behaviors. *International Journal of Behavioral Development*, *35*(2), 116-124.

Carlo, G., Koller, S. H., Eisenberg, N., Da Silva, M. S. & Frohlich, C. B. (1996). A cross-national study on the relations among prosocial moral reasoning, gender role orientations, and prosocial behaviors. *Developmental Psychology*, *32*(2), 231.

Caspi, A., Harrington, H., Milne, B., Amell, J. W., Theodore, R. F. & Moffitt, T. E. (2003). Children's behavioral styles at age 3 are linked to their adult personality traits at age 26. *Journal of Personality*, *71*(4), 495-514.

CBS (2017). *Bevolking: generatie, geslacht, leeftijd en herkomstgroepering, 1 januari*. Retrieved from http://statline.cbs.nl/statweb/publication/?vw=t&dm=slnl&pa=37325&d1=a&d2=0&d3=0&d4=0&d5=0-4,137,152,220,237&d6=0,4,9,14,18-19&hd=151214-1201&hdr=g2,g1,g3,t&stb=g4,g5D

Cheung, G. W. & Rensvold, R.B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance, *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233-255.

CITO (2017). *Toetsen op School Primair Onderwijs.* Arnhem: CITO.

Coie, J. D. & Dodge, K. A. (1983). Continuities and changes in children's social status: A five-year longitudinal study. *Merrill-Palmer Quarterly*, 261-282.

Cook, C., Williams, K. R., Guerra, N. G., Kim, T. & Sadek, S. (2010). Predictors of childhood bullying and victimization. *School Psychology Quarterly, 25(2), 65–83.*

Crothers, L. M. & Levinson, E. M. (2004). Assessment of bullying: A review of methods and instruments. *Journal of Counseling & Development*, *82*(4), 496-503.

Cuevas, A. K. (2011). *The Effects of Perceived Peer Behavior and Visual Art Content on Students' Moral Action Confidence.* Retrieved from: http://www.eric.ed.gov/PDFS/ED518587.pdf

Cumberland-Li, A., Eisenberg, N. & Reiser, M. (2004). Relations of young children's agreeableness and resiliency to effortful control and impulsivity. *Social Development,* 13(2), 193-212.

Dake, J. A., Price, J. H. & Telljohann, S. K. (2003). The nature and extent of bullying at school. *Journal of School Health, 73*(5), 173-180.

Davidson, P., Turiel, E. & Black, A. (1983). The effect of stimulus familiarity on the use of criteria and justifications in children's social reasoning. *British Journal of Developmental Psychology, 1*(1), 49-65.

Davis, H. A. (2003). Conceptualizing the role and influence of student-teacher relationships on children's social and cognitive development. *Educational Psychologist*, *38*(4), 207-234.

De Bolle, M. & Tackett, J. L. (2013). Anchoring Bullying and Victimization in Children Within a Five Factor Model Based Person Centred Framework. *European Journal of Personality, 27*(3), 280-289.

Deković, M. & Janssens, J. M. (1992). Parents' child-rearing style and child's sociometric status. *Developmental Psychology*, *28*(5), 925.

Dodge, K. A. (2006). Translational science in action: Hostile attributional style and the development of aggressive behavior problems. *Development and Psychopathology*, *18*(3), 791-814.

Dodge, K. A., Greenberg, M. T. & Malone, P. S. (2008). Testing an idealized dynamic cascade model of the development of serious violence in adolescence. *Child Development*, *79*(6), 1907-1927.

Dodge, K. A. & Sherrill, M. R. (2006). Deviant peer group effects in youth mental health interventions. *Deviant Peer Influences in Programs for Youth: Problems and Solutions*, 97-121.

Domitrovich, C. E., Durlak, J. A., Staley, K. C. & Weissberg, R. P. (2017). Social-Emotional Competence: An Essential Factor for Promoting Positive Adjustment and Reducing Risk in School Children. *Child Development*, *88*(2), 408-416.

Donnellan, M. B., Trzesniewski, K. H., Robins, R. W., Moffitt, T. E. & Caspi, A. (2005). Low self-esteem is related to aggression, antisocial behavior, and delinquency. *Psychological Science*, *16*(4), 328-335.

Downey, J L.A.. Mountstephen, J. L., Hansen, K., & Stough, C. (2008). Emotional intelligence and scholastic achievement in Australian adolescents *Australian Journal of Psychology, 60,* 10–17.

Driessen, G., Mulder, L., Ledoux, G., Roeleveld, J. & Veen, I. V. (2009). *Cohortonderzoek COOL 5-18: technisch rapport basisonderwijs, eerste meting 2007/08.* Nijmegen/Amsterdam: ITS / SCO-Kohnstamm Instituut.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*(3-4), 169-200.

Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annual Review of Psychology*, *51*(1), 665-697.

Eisenberg, N., Carlo, G., Murphy, B. & Court, P. (1995). Prosocial development in late adolescence: a longitudinal study. *Child development*, *66*(4), 1179-1197.

Eisenberg, N. & Fabes, R. A. (1998). Prosocial development. In W. Damon (Series Ed.) & N. Eisenberg (Volume Ed.), *Handbook of Child Psychology: Vol. 3 Social, Emotional, and Personality Development* (pp. 701-778). New York NY: John Wiley & Sons, Inc.

Eisenberg, N., Fabes, R. A., Murphy, B., Karbon, M., Smith, M. & Maszk, P. (1996). The relations of children's dispositional empathy-related responding to their emotionality, regulation, and social functioning. *Developmental Psychology, 32*(2), 195.

Eisenberg, N., Fabes, R. A. & Spinrad, T. L. (2006). Prosocial behavior. In N. Eisenberg (Vol. Ed.) and W. Damon & R.M. Lerner (Series Eds.), *Handbook of Child Psychology: Vol. 3 Social, Emotional, and Personality Development* (6th ed.; pp. 646-718). Hoboken, NJ: John Wiley & Sons.

Eisenberg, N., Guthrie, I. K., Cumberland, A., Murphy, B. C., Shepard, S. A., Zhou, Q. & Carlo, G. (2002). Prosocial development in early adulthood: a longitudinal study. *Journal of Personality and Social Psychology*, *82*(6), 993.

Eisenberg, N., Shea, C. L., Carlo, G. & Knight, G. P. (2014). Empathy-related responding and cognition: A "chicken and the egg" dilemma. In W.M. Kurtines & J.L. Gewirtz (Eds.) *Handbook of Moral Behavior and Development* (Vol. 2; pp. 63-88). New York: Psychology Press.

Eisenberg, N., Spinrad, T. L. & Eggum, N. D. (2010). Emotion-related self-regulation and its relation to children's maladjustment. *Annual Review of Clinical Psychology*, *6*, 495-525.

Eisenberg, N., Spinrad, T. L. & Knafo-Noam, A. (2015). Prosocial development. In M. E. Lamb & C.G. Coll (Eds.), *Handbook of Child Psychology and Developmental Science, Vol. 3: Socioemotional Processes* (pp. 610-656). New York: Wiley.

Eisenberg, N., Wentzel, M. & Harris, J. D. (1998). The role of emotionality and regulation in empathy-related responding. *School Psychology Review, 27*(4), 506.

Ellis, L. K. & Rothbart, M. K. (1999). *Early adolescent temperament questionnaire-revised*. Oregon: University of Oregon.

Espelage, D. L., Bosworth, K. & Simon, T. R. (2001). Short-term stability and prospective correlates of bullying in middle-school students: An examination of potential demographic, psychosocial, and environmental influences. *Violence and Victims, 16*(4), 411-426.

Espelage, D. L., Holt, M. K. & Henkel, R. R. (2003). Examination of peer-group contextual effects on aggression during early adolescence. *Child Development, 74*(1), 205-220.

Fabes, R. A., Carlo, G., Kupanoff, K. & Laible, D. (1999). Early adolescence and prosocial/moral behavior I: The role of individual processes. *The Journal of Early Adolescence, 19*(1), 5-16.

Fabes, R. A., Martin, C. L. & Hanish, L. D. (2002). *The role of sex segregation in young children's prosocial behavior and disposition.* In Groningen Conference on Prosocial Dispositions and Solidarity, Groningen, The Netherlands.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C. & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272-299.

Faris, R. & Ennett, S. (2012). Adolescent aggression: The role of peer group status motives, peer aggression, and group characteristics. *Social Networks, 34*(4), 371-378.

Fink, K. & Slade, S. (Eds.). (2016). *Integrating Prosocial Learning with Education Standards: School Climate Reform Initiatives.* New York: Taylor & Francis.

Fossati, A., Borroni, S. & Maffei, C. (2012). Bullying as a style of personal relating: personality characteristics and interpersonal aspects of self-reports of bullying behaviours among Italian adolescent high school students. *Personality and Mental Health, 6*(4), 325-339.

Frey, K. S., Hirschstein, M. K., Edstrom, L. V. & Snell, J. L. (2009). Observed reductions in school bullying, nonbullying aggression, and destructive bystander behavior: A longitudinal evaluation. *Journal of Educational Psychology*, *101*(2), 466.

Gasser, L. Gutzwiller-Helfenfinger, E., Latzko, B. & Malti. T. (2013). Moral emotion attributions and moral motivation. In K. Heinrichs, F. Oser &T. Lovat (Eds.), *Handbook of Moral Motivation* (pp. 267-287). Rotterdam: SensePublishers.

Gasser, L. & Keller, M. (2009). Are the Competent the Morally Good? Perspective Taking and Moral Motivation of Children Involved in Bullying. *Social Development, 18*(4), 798-816.

Gasser, L., Malti, T. & Gutzwiller-Helfenfinger, E. (2012). Aggressive and nonaggressive children's moral judgments and moral emotion attributions in situations involving retaliation and unprovoked aggression. *The Journal of Genetic Psychology, 173*(4), 417-439.

Gewirth, A. (1984). The epistemology of human rights. *Social Philosophy and Policy, 1*(2), 1-24.

Gini, G. (2006). Social cognition and moral cognition in bullying: What's wrong?. *Aggressive Behavior, 32*(6), 528-539.

Gini, G., Albiero, P., Benelli, B. & Altoè, G. (2007). Does empathy predict adolescents' bullying and defending behavior? *Aggressive Behavior, 33*(5), 467-476.

Gini, G., Albiero, P., Benelli, B. & Altoe, G. (2008). Determinants of adolescents' active defending and passive bystanding behavior in bullying. *Journal of Adolescence, 31*(1), 93-105.

Gini, G., Pozzoli, T. & Bussey, K. (2014). Collective moral disengagement: Initial validation of a scale for adolescents. *European Journal of Developmental Psychology, 11*(3), 386-395.

Gini, G., Pozzoli, T. & Bussey, K. (2015). The role of individual and collective moral disengagement in peer aggression and bystanding: a multilevel analysis. *Journal of Abnormal Child Psychology, 43*(3), 441-452.

Gini, G., Pozzoli, J. & Hauser, M. (2011). Bullies have enhanced moral competence to judge relative to victims, but lack moral compassion. *Personality and Individual Differences, 50*(5), 603–608.

GION, Rijksuniversiteit Groningen; CITO Arnhem (2008): *Cohortonderzoek Onderwijsloopbanen van 5-18 jaar - COOL 5-18 - Voortgezet Onderwijs Klas 3 - 2007/2008*. DANS. https://doi.org/10.17026/dans-zgw-entf

Goldberg, L. R. (2001). Analyses of Digman's child-personality data: Derivation of Big-Five factor scores from each of six samples. *Journal of Personality*, *69*(5), 709-744.

Goossens, F. A., Olthof, T. & Dekker, P. H. (2006). New participant role scales: Comparison between various criteria for assigning roles and indications for their validity. *Aggressive Behavior*, *32*(4), 343-357.

Gorsuch, R. L. (1990). Common factor analysis versus component analysis: Some well and little known facts. *Multivariate Behavioral Research*, *25*(1), 33-39.

Gravemeijer, K. P. E. & Kirschner, P. A. (2007). Naar meer evidence-based onderwijs? *Pedagogische Studiën, 84*(6), 463-472.

Graziano, W. G. & Eisenberg, N. H. (1997). Agreeableness: A dimension of personality. In R. Hogan, J. Johnson & S. Briggs (Eds.), *Handbook of Personality Psychology* (pp. 795-824). San Diego, CA: Academic Press.

Greenberg, M. T., Kusche, C. A., Cook, E. T. & Quamma, J. P. (1995). Promoting emotional competence in school-aged children: The effects of the PATHS curriculum. *Development and Psychopathology*, *7*(1), 117-136.

Grusec, J. E. & Redler, E. (1980). Attribution, reinforcement, and altruism: A developmental analysis. *Developmental Psychology*, *16*(5), 525.

Guerra, N. G., Nucci, L. & Huesmann, L. R. (1994). Moral cognition and childhood aggression. In *Aggressive Behavior* (pp. 13-33). Springer US.

Gutzwiller-Helfenfinger, E. (2008). Die Wirkung von erweitertem Rollenspiel auf soziale Perspektivenübernahme und antisoziales Verhalten. In T. Malti & S. Perren (Hrsg), *Entwicklung und Förderung sozialer Kompetenzen in Kindheit und Adoleszenz* (pp. 231-245). Stuttgart: Kohlhammer-Verlag.

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814.

Halverson, C. F., Havill, V. L., Deal, J., Baker, S. R., Victor, J. B., Pavlopoulos, V., ... & Wen, L. (2003). Personality structure as derived from parental ratings of free descriptions of children: The Inventory of Child Individual Differences. *Journal of Personality*, *71*(6), 995-1026.

Hampson, S. E., Goldberg, L. R., Vogt, T. M. & Dubanoski, J. P. (2007). Mechanisms by which childhood personality traits influence adult health status: educational attainment and healthy behaviors. *Health Psychology*, *26*(1), 121.

Hardy, S. (2006). Identity, Reasoning, and Emotion: An Empirical Comparison of Three Sources of Moral Motivation. *Motivation and Emotion, 30* (3), 205-213.

Hawker, D. S. & Boulton, M. J. (2000). Twenty years' research on peer victimization and psychosocial maladjustment: a meta-analytic review of cross-sectional studies. *Journal of Child Psychology and Psychiatry, 41*(4), 441-455.

Hawkins, Pepler & Craig (2001). Naturalistic observations of peer interventions in bullying. *Social Development, 10*(4), 512-527.

Hawley, P. H., Little, T. D. & Card, N. A. (2007). The allure of a mean friend: Relationship quality and processes of aggressive adolescents with prosocial skills. *International Journal of Behavioral Development*, *31*(2), 170-180.

Helwig, C. C., Tisak, M. S. & Turiel, E. (1990). Children's social reasoning in context: Reply to Gabennesch. *Child Development, 61*(6), 2068-2078.

Hendriks, A. A. J. (1997). *The construction of the five-factor personality inventory (FFPI).* Unpublished doctoral dissertation. Groningen: University of Groningen.

Hepach, R., Vaish, A., Grossmann, T. & Tomasello, M. (2016). Young children want to see

others get the help they need. *Child Development, 87*(6), 1703-1714.

Hoffman, M. L. (2008). Empathy and Prosocial Behavior. In M. Lewis, J. M. Haviland-Jones & L. F. Barrett (Eds.), *Handbook of Emotions* (pp. 440-455). New York: Guilford Press.

Hoogeveen, P. & Winkels, J. (2008). *Het Didactische Werkvormenboek.* Assen: Uitgeverij Van Gorcum.

Hughes, C. & Leekam, S. (2004). What are the links between theory of mind and social relations? Review, reflections and new directions for studies of typical and atypical development. *Social Development, 13*(4), 590-619.

Huitsing, G. (2014). *A social network perspective on bullying* (Doctoral Dissertation). Groningen: University of Groningen.

Huitsing, G. & Veenstra, R. (2012). Bullying in classrooms: Participant roles from a social network perspective. *Aggressive Behavior, 38,* 494-509.

Hymel, S., Rocke-Henderson, N. & Bonanno, R.A. (2005). Moral disengagement: A framework for understanding bullying among adolescents.. *Journal of Social Sciences*, *8*, 1–11.

ITS & Kohnstamm Instituut (2009). *Cohortonderzoek Onderwijsloopbanen van 5-18 jaar - COOL 5-18 - Basisonderwijs 2007/08.* [Data set] Retrieved from: https://easy.dans.knaw.nl/

Jansen, P. W., Verlinden, M., Dommisse-van Berkel, A., Mieloo, C., van der Ende, J., Veenstra, R., ... & Tiemeier, H. (2012). Prevalence of bullying and victimization among children in early elementary school: Do family and school neighbourhood socioeconomic status matter?. *BMC Public Health, 12*(1), 1.

Jansma, D.J., Malti, T., Opdenakker, M.C. J. L. & Van der Werf, M. P. C. (in press). Assessment of Anticipated Emotions following Moral Transgressions. *European Journal of Psychological Assessment.*

Jansma, D.J., Malti, T., Opdenakker, M.C.J.L. & Van der Werf, M.P.C. (2017). *Individual and Class Moral Correlates of Children's Bullying.* Manuscript submitted for publication.

Jennings, P. A. & Greenberg, M. T. (2009). The prosocial classroom: Teacher social and emotional competence in relation to student and classroom outcomes. *Review of Educational Research, 79*(1), 491-525.

Jensen-Campbell, L. A. & Graziano, W. G. (2001). Agreeableness as a moderator of interpersonal conflict. *Journal of Personality*, *69*(2), 323-362.

Jensen-Campbell, L. A., Gleason, K. A., Adams, R. & Malcolm, K. T. (2003). Interpersonal conflict, agreeableness, and personality development. *Journal of Personality*, *71*(6), 1059-1086.

Jensen-Campbell, L. A. & Malcolm, K. T. (2007). The importance of conscientiousness in adolescent interpersonal relationships. *Personality and Social Psychology Bulletin, 33*(3), 368-383.

Johnson, D.W. & Johnson, R.T. (1999). Making cooperative learning work. *Theory into Practice, 38*(2), 67-73.

Jolliffe, D. & Farrington, D. P. (2006). Development and validation of the Basic Empathy Scale. *Journal of Adolescence, 29*(4), 589-611.

Jones, L. and Jones, V. (2010). *Comprehensive Classroom Management, creating communities of support and solving problems.* New Jersey: Merill.

Jones, S. E., Miller, J. D. & Lynam, D. R. (2011). Personality, antisocial behavior, and aggression: A meta-analytic review. *Journal of Criminal Justice*, *39*(4), 329-337.

Jordan J. (2007). Taking the first step toward a moral action*:* a review of moral sensitivity measurement across domains. *The Journal of Genetic Psychology, 168* (3), 323–9.

Juvonen, J. & Galvan, A. (2008). Peer influence in involuntary social groups: Lessons from research on bullying. In: M. J. Prinstein & K. A. Dodge (Eds.) *Understanding Peer Influence in Children and Adolescents* (pp. 225-244). New York: Guilford Press.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*(1), 141-151.

Kaukiainen, A., Björkqvist, K., Lagerspetz, K., Österman, K., Salmivalli, C., Rothberg, S. & Ahlbom, A. (1999). The relationships between social intelligence, empathy, and three types of aggression. *Aggressive Behavior, 25*(2), 81-89.

Kaukiainen, A., Björkqvist, K., Österman, K. & Lagerspetz, K. M. (1996). Social intelligence and empathy as antecedents of different types of aggression. *Annals of the New York Academy of Sciences, 794*(1), 364-366.

Keller, M., Lourenço, O., Malti, T. & Saalbach, H. (2003). The multifaceted phenomenon of 'happy victimizers': a cross-cultural comparison of moral emotions. *British Journal of Developmental Psychology*, *21*(1), 1-18.

Koenig, A. L., Cicchetti, D. & Rogosch, F. A. (2004). Moral development: The association between maltreatment and young children's prosocial behaviors and moral transgressions. *Social Development*, *13*(1), 87-106.

Kohlberg, L. (1969). *Stage* and sequence: The cognitive-developmental approach to socialization. In D.A. Goslin (Ed.), *Handbook of Socialization Theory and Research* (pp. 347-480). Chicago: Rand McNally.

Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M. & Doolaard, S. (2016). A meta-analysis of the effects of classroom management strategies and classroom management programs on students' academic, behavioral, emotional, and motivational outcomes. *Review of Educational Research*, *86*(3), 643-680.

Krettenauer, T. & Eichler, D. (2006). Adolescents' self-attributed moral emotions following a moral transgression: Relations with delinquency, confidence in moral judgment and age. *British Journal of Developmental Psychology*, *24*(3), 489-506.

Krettenauer, T., Malti, T. & Sokol, B. W. (2008). The development of moral emotion expectancies and the happy victimizer phenomenon: A critical review of theory and application. *International Journal of Developmental Science*, *2*(3), 221-235.

Krueger, R. F., Hicks, B. M. & McGue, M. (2001). Altruism and antisocial behavior: Independent tendencies, unique personality correlates, distinct etiologies. *Psychological Science*, *12*(5), 397-402.

Laible, D., Carlo, G., Panfile, T., Eye, J. & Parker, J. (2010). Negative emotionality and emotion regulation: A person-centered approach to predicting socioemotional adjustment in young adolescents. *Journal of Research in Personality*, *44*(5), 621-629.

Lapsley, D. K., & Hill, P. L. (2009). The development of the moral personality. In D. Narvaez & D. Lapsley (Eds.), *Personality, Identity, and Character: Explorations in Moral Psychology* (pp.185–213). New York: Cambridge University Press.

Lemerise, E. A. & Arsenio, W. F. (2000). An integrated model of emotion processes and cognition in social information processing. *Child Development*, *71*(1), 107-118.

Little, T. D., Lindenberger, U. & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When" good" indicators are bad and" bad" indicators are good. *Psychological Methods, 4*(2), 192-211.

Long, J. (1997). *Regression models for categorical and limited dependent variables.* Thousand Oaks: Sage.

Lourenço, O. (1997). Children's attributions of moral emotions to victimizers: Some data, doubts and suggestions. *British Journal of Developmental Psychology*, *15*(4), 425-438.

Malti, T. (2016). Toward an integrated clinical-developmental model of guilt. *Developmental Review, 39*, 16-36.

Malti, T., Chaparro, M. P., Zuffianò, A. & Colasante, T. (2016). School-based interventions to promote empathy-related responding in children and adolescents: A developmental analysis. *Journal of Clinical Child & Adolescent Psychology*, *45*(6), 718-731.

Malti, T. & Dys, S.P. (in press). From being nice to being kind: Development of prosocial behaviors. *Current Opinion in Psychology.*

Malti, T., Dys, S.P., Colasante, T. & Peplak, J. (in press). Emotions and morality: New developmental perspectives. In C. Helwig (Vol. Ed.) and M. Harris (Series Ed.), *Current Issues in Developmental Psychology: New Perspectives on Moral Development.* New York: Psychology Press.

Malti, T., Gasser, L. & Buchmann, M. (2009). Aggressive and prosocial children's emotion attributions and moral reasoning. *Aggressive Behavior*, *35*(1), 90-102.

Malti, T., Gummerum, M., Keller, M. & Buchmann, M. (2009). Children's moral motivation, sympathy, and prosocial behavior. *Child Development*, *80*(2), 442-460.

Malti, T. & Keller, M. (2010). Development of moral emotions in cultural context. In W. Arsenio & E. Lemerise (Eds.), *Emotions, Aggression, and Morality in Children: Bridging Development and Psychpathology* (pp. 177–198). Washington, DC: American Psychological Association.

Malti, T., Killen, M. & Gasser, L. (2012). Social judgments and emotion attributions about exclusion in Switzerland. *Child Development, 83*(2), 697-711.

Malti, T. & Krettenauer, T. (2013). The relation of moral emotion attributions to prosocial and antisocial behavior: A meta-analysis. *Child Development*, *84*(2), 397-412.

Malti, T. & Latzko, B. (2012). Moral emotions. *Encyclopedia of Human Behavior, 2*, 644-649.

Malti, T. & Ongley, S.F. (2014). The development of moral emotions and moral reasoning. In M. Killen & J. Smetana (Eds.), *Handbook of Moral Development* (pp.163-183). New York: Taylor & Francis.

Malti, T., Ongley, S. F., Dys, S. P. & Colasante, T. (2012). Adolescents' emotions and reasoning in contexts of moral conflict and social exclusion. *New Directions for Student Leadership, 2012*(136), 27-40.

Malti, T., Ongley, S. F., Peplak, J., Chaparro, M. P., Buchmann, M., Zuffianò, A. & Cui, L. (2016). Children's Sympathy, Guilt, and Moral Reasoning in Helping, Cooperation, and Sharing: A 6-Year Longitudinal Study. *Child Development, 87*(6), 1783-1795.

Mascolo, M. & Fischer, K. (2010). The dynamic development of thinking, feeling, and acting over the life span. In W. F. Overton & R. M. Lerner (Eds.), *The Handbook of Life-span Development* (Vol. 1, pp. 149–194). Hoboken, NJ: Wiley.

McAdams. D.P. & Adler, J.M. (2006). How Does Personality Develop? In: D.K. Mroczek & T.D. Little (Eds.) *Handbook of Personality Development* (pp. 469-492). Mahwah, N.J.: Lawrence Erlbaum.

McCrae, R. R. & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*(2), 175-215.

Meijer, G. (2012). *Instrument om morele sensitiviteit bij kinderen te meten. Een studie naar empathie, sociaal perspectief nemen/ Theory of Mind (ToM) in situaties die toegespitst zijn op pesten.* [Master's thesis, University of Groningen].

Menesini, E. & Camodeca, M. (2008). Shame and guilt as behaviour regulators: Relationships with bullying, victimization and prosocial behaviour. *British Journal of Developmental Psychology, 26*(2), 183-196.

Menesini, E., Camodeca, M. & Nocentini, A. (2010). Bullying among siblings: The role of personality and relational variables. *British Journal of Developmental Psychology, 28*(4), 921-939.

Menesini, E., Sanchez, V., Fonzi, A., Ortega, R., Costabile, A. & Lo Feudo, G. (2003). Moral emotions and bullying: A cross-national comparison of differences between bullies, victims and outsiders. *Aggressive Behavior, 29*(6), 515-530.

Merrell, K.W., Gueldner, B, A., Ross, S. W. & Isava, D. M. (2008). How effective are school bullying intervention programs? A meta-analysis of intervention research. *School Psychology Quarterly, 23*(1), 26-42.

Miller, J. G. (2006). Insights into moral development from cultural psychology. In M. Killen & J. Smetana (Eds.), *Handbook of Moral Development* (pp. 375-398). New York: Taylor & Francis.

Miller, J., Flory, K., Lynam, D. & Leukefeld, C. (2003). A test of the four-factor model of impulsivity-related traits. *Personality and Individual Differences, 34*(8), 1403-1418.

Miller, J. D. & Lynam, D. (2001). Structural models of personality and their relation to antisocial behavior: A meta-analytic review. *Criminology*, *39*(4), 765-798.

Montada, L. (1993). Understanding oughts by assessing moral reasoning or moral emotions. In G. Noam & T. Wren (Eds.), *The Moral Self* (pp. 292-309). Boston: MIT-Press.

Morton K. R, Worthley, J. S., Testerman, J. K & Mahoney, M.L. (2006). Defining features of moral sensitivity and moral motivation: pathways to moral reasoning in medical students. *Journal of Moral Education, 35*(3), 387-406.

Mower, D., Robison, W. L. & Vandenberg, P. (2015). *Developing Moral Sensitivity.* New York: Routledge.

Mroczek, D. K. & Little, T. D. (2014). *Handbook of Personality Development*. New York: Psychology Press.

Muthén, L.K. and Muthén, B.O. (1998-2015). *Mplus User's Guide. Seventh Edition.* Los Angeles, CA: Muthén & Muthén.

Myyry, L., Juujärvi, S. & Pesso, K. (2010). Empathy, perspective taking and personal values as predictors of moral schemas. *Journal of Moral Education, 39* (2), 213–233.

Nansel, T. R., Craig, W., Overpeck, M. D., Saluja, G. & Ruan, W. J. (2004). Cross-national consistency in the relationship between bullying behaviors and psychosocial adjustment. *Archives of Pediatrics and Adolescent Medicine, 158*(8), 730-736.

Nantel-Vivier, A., Kokko, K., Caprara, G. V., Pastorelli, C., Gerbino, M. G., Paciello, M., ...& Tremblay, R. E. (2009). Prosocial development from childhood to adolescence: a multi-informant perspective with Canadian and Italian longitudinal studies. *Journal of Child Psychology and Psychiatry, 50*(5), 590-598.

Narvaez, D. & Nucci, L. P. (Eds.). (2008). *Handbook of Moral and Character Education* (pp. 310-327). Routledge.

Narvaez, D. & Vaydich, J. (2008). Moral development and behaviour under the spotlight of the neurobiological sciences. *Journal of Moral Education, 37*(3), 289-313.

Neal, J. W. (2008). "Kracking" the missing data problem: applying Krackhardt's cognitive social structures to school-based social networks. *Sociology of Education, 81*(2), 140-162.

Newcomb, A. F., Bukowski, W. M. & Pattee, L. (1993). Children's peer relations: a meta-analytic review of popular, rejected, neglected, controversial, and average sociometric status. *Psychological Bulletin, 113*(1), 99-128.

Nickerson, A. B., Mele, D. & Princiotta, D. (2008). Attachment and empathy as predictors of roles as defenders or outsiders in bullying interactions. *Journal of School Psychology, 46*(6), 687-703.

Nishina, A., Juvonen, J. & Witkow, M. R. (2005). Sticks and stones may break my bones, but names will make me feel sick: The psychosocial, somatic, and scholastic consequences of peer harassment. *Journal of Clinical Child and Adolescent Psychology, 34*(1), 37-48.

NJI (2015). *Beoordeling anti-pestprogramma's.* Utrecht: Nederlands Jeugdinstituut.

Nucci, L. (1981). Conceptions of personal issues: A domain distinct from moral or societal concepts. *Child Development, 52*(1), 114-121.

Nunner-Winkler, G. (1999). Development of moral understanding and moral motivation. In F. E Weinert & W. Schneider, (Eds.), *Individual development from 3 to 12. Findings from the Munich Longitudinal Study* (pp. 253–290). New York: Cambridge University Press.

Nunner-Winkler, G. (2007). Development of moral motivation from childhood to early adulthood. *Journal of Moral Education, 36*(4), 399-414.

Nunner-Winkler, G. (2008). From super-ego and conformist habitus to ego-syntonic moral motivation. Sociohistoric changes in moral motivation. *International Journal of Developmental Science*, *2*(3), 251-268.

Nunner-Winkler, G. (2013). Moral motivation and the happy victimizer phenomenon. In K. Heinrichs, F. Oser &T. Lovat (Eds.), *Handbook of Moral Motivation* (pp. 267-287). Rotterdam: SensePublishers.

Nunner-Winkler, G. & Sodian, B. (1988). Children's understanding of moral emotions. *Child Development*, *59*(5), 1323-1338.

Nunnally, J.C. & Bernstein, I.H. (1994). *Psychometric Theory* (3rd ed.). New York: McGras-Hill.

Nucci, L. P. (2001). *Education in the moral domain.* Cambridge: Cambridge University Press.

Oberle, E. & Schonert-Reichl, K. A. (2013). Relations among peer acceptance, inhibitory control, and math achievement in early adolescence. *Journal of Applied Developmental Psychology, 34*(1), 45-51.

Olthof, T. (2010). Conscience in the classroom: Moral emotions, moral judgments, and moral identity as predictors of children's pro-social and antisocial interpersonal behavior. In W. Koops, D. Brugman, T.J. Ferguson & A.F. Sanders (Eds.), *The Development and Structure of Conscience* (pp. 327-341). New York: Psychology Press.

Olthof, T., Goossens, F. A., Vermande, M. M., Aleva, E. A. & van der Meulen, M. (2011). Bullying as strategic behavior: Relations with desired and acquired dominance in the peer group. *Journal of School Psychology, 49*(3), 339-359.

Olthof, T., Schouten, A., Kuiper, H., Stegge, H. & Jennekens-Schinkel, A. (2000). Shame and guilt in children: Differential situational antecedents and experiential correlates. *British Journal of Developmental Psychology, 18*(1), 51-64.

Olweus, D. (1993). Victimization by peers: Antecedents and long-term outcomes. Social withdrawal, inhibition, and shyness in childhood. In: K. H. Rubin & J. B. Asendorpf (Eds.) *Social Withdrawal, Inhibition, and Shyness in Childhood* (pp. 315-341). New York: Lawrence Erlbaum Associates, Inc.

Padilla-Walker, L. M. & Christensen, K. J. (2011). Empathy and self-regulation as mediators between parenting and adolescents' prosocial behavior toward strangers, friends, and family. *Journal of Research on Adolescence, 21*(3), 545-551.

Peabody, D. & De Raad, B. (2002). The substantive nature of psycholexical personality factors: a comparison across languages. *Journal of Personality and Social Psychology*, *83*(4), 983.

Pekrun, R. (2016). Academic emotions. In K. R. Wentzel & D.B. Miele (Eds.), *Handbook of Motivation at School* (pp. 120-144). New York: Routledge.

Pellegrini, A. D. & Long, J. D. (2002). A longitudinal study of bullying, dominance, and

victimization during the transition from primary school through secondary school. *British Journal of Developmental Psychology, 20*(2), 259-280.

Perren, S. & Alsaker, F. D. (2006). Social behavior and peer relationships of victims, bully-victims, and bullies in kindergarten. *Journal of Child Psychology and Psychiatry, 47*(1),45-57.

Perren, S. & Gutzwiller-Helfenfinger, E. (2012). Cyberbullying and traditional bullying in adolescence: Differential roles of moral disengagement, moral emotions, and moral values. *European Journal of Developmental Psychology, 9*(2), 195-209.

Perren, S., Gutzwiller-Helfenfinger, E., Malti, T. & Hymel, S. (2012). Moral reasoning and emotion attributions of adolescent bullies, victims, and bully-victims. *British Journal of Developmental Psychology, 30*(4), 511-530.

Piaget, J. (1965). *Les Études Sociologiques*. Genève: Droz.

Pozzoli, T. & Gini, G. (2010). Active defending and passive bystanding behavior in bullying: The role of personal characteristics and perceived peer pressure. *Journal of Abnormal Child Psychology, 38*(6), 815-827.

Pozzoli, T. & Gini, G. (2012). Why do bystanders of bullying help or not? A multidimensional model. *The Journal of Early Adolescence, 33*(3), 315-340.

Pozzoli, T., Gini, G. & Vieno, A. (2012). The role of individual correlates and class norms in defending and passive bystanding behavior in bullying: A multilevel analysis. *Child Development, 83*(6), 1917-1931.

Putnam, S. P., Ellis, L. K. & Rothbart, M. K. (2001). The structure of temperament from infancy through adolescence. *Advances in Research on Temperament*, 165-182.

Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., ... & Lewis, T. (2000). *A User's Guide to MLwiN.* London: Institute of Education.

Rasbash, J., Charlton, C., Browne, W.J., Healy, M. & Cameron, B. (2005). *MLwiN Version 2.02.* Centre for Multilevel Modelling: University of Bristol.

Reiman, A.J. & Dotger, B.H. (2008). What does innovation mean for moral educators? *Journal of Moral Education, 37*(2), 151-164.

Rest, J. R. (1983). Morality, in: J. H. Flavell & E. M. Markman (Eds), *Handbook of Child Psychology. Vol. III: Cognitive Development* (4th edn), New York: John Wiley, 556–629.

Rest, J. (1986). *Moral Development: Advances in Research and Theory.* New York: Praeger.

Rest, J. R. (1994). *Moral Development in the Professions: Psychology and Applied Ethics.* New York: Psychology Press.

Rivara, F. & Le Menestrel, S. (2016*). Preventing Bullying Through Science, Policy, and Practice*. Washington (DC): National Academies Press.

Robins, R.W., John, O.P. & Caspi, A. (1994). Major dimensions of personality in early adolescence: The Big Five and beyond. In C.F. Halverson, G.A. Kohnstamm & R.P. Martin (Eds.), *The Developing Structure of Temperament and Personality from Infancy to Adulthood* (pp. 267–291). Hillsdale, NJ: Erlbaum.

Rothbart, M. K., Derryberry, D. & Posner, M. I. (1994). A psychobiological approach to the development of temperament. *Temperament: Individual Differences at the Interface of Biology and Behavior*, 83-116.

Rothbart, M.K. & Bates, J.E. (1998). Temperament. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of Child Psychology: Vol. 3. Social, Emotional, and Personality Development* (5th ed., pp. 105-76). New York: Wiley & Sons.

Rothbart, M. & Park, B. (1986). On the confirmability and disconfirmability of trait concepts. *Journal of Personality and Social Psychology*, *50*(1), 131.

Rubin, K. H., Bukowski, W. M. & Laursen, B. (Eds.). (2011). *Handbook of Peer Interactions, Relationships, and Groups.* New York: Guilford Press.

Rubin, K. H., Bukowski, W. M. & Parker, J. G. (1998). *Peer Interactions, Relationships, and Groups. Handbook of Child Psychology*. New York: John Wiley & Sons.

Rupp, J. C.C. & Veugelers, W. (2003). *Moreel-politieke heroriëntatie in het onderwijs.* Leuven: Garant.

Salmivalli, C. (2010). Bullying and the peer group: A review. *Aggression and Violent Behavior, 15*(2), 112-120.

Salmivalli, C. & Isaacs, J. (2005). Prospective relations among victimization, rejection, friendlessness, and children's self-and peer-perceptions. *Child Development, 76*(6), 1161-1171.

Salmivalli, C., Lagerspetz, K., Björkqvist, K., Österman, K. & Kaukiainen, A. (1996). Bullying as a group process: Participant roles and their relations to social status within the group. *Aggressive Behavior, 22*(1), 1-15.

Salmivalli, C., Lappalainen, M. & Lagerspetz, K. M. (1998). Stability and change of behavior in connection with bullying in schools: A two-year follow-up. *Aggressive Behavior, 24*(3), 205-218.

Salmivalli, C. & Voeten, M. (2004). Connections between attitudes, group norms, and behaviour in bullying situations. *International Journal of Behavioral Development, 28*(3), 246-258.

Scheithauer, H., Hayer, T., Petermann, F. & Jugert, G. (2006). Physical, verbal, and relational forms of bullying among German students: Age trends, gender differences, and correlates. *Aggressive Behavior, 32*(3), 261-275.

Schwarz, N. & Clore, G. L. (1996). Feelings and phenomenal experiences. In E. T. Higgins & A. Kruglanski (Eds.), *Social Psychology: Handbook of Basic Principles* (pp. 433–465). New York: Guilford.

Seals, D. & Young, J. (2003). Bullying and victimization: Prevalence and relationship to gender, grade level, ethnicity, self-esteem, and depression. *Adolescence, 38*(152), 735.

Shiner, R. (2006). Temperament and personality in childhood. In: D.K. Mroczek & T.D. Little (Eds.) *Handbook of Personality Development* (pp. 213-230). Mahwah, N.J.: Lawrence Erlbaum.

Shiner, R. & Caspi, A. (2003). Personality differences in childhood and adolescence: Measurement, development, and consequences. *Journal of Child Psychology and Psychiatry*, *44*(1), 2-32.

Shiner, R. L. & Masten, A. S. (2008). Personality in childhood: A bridge from early temperament to adult outcomes. *International Journal of Developmental Science*, *2*(1-2), 158-175.

Shiner, R. L., Masten, A. S. & Tellegen, A. (2002). A developmental perspective on personality in emerging adulthood: Childhood antecedents and concurrent adaptation. *Journal of Personality and Social Psychology*, *83*(5), 1165-1177.

Singer, D. G., Golinkoff, R. M. & Hirsh-Pasek, K. (2006). *Play= Learning: How Play Motivates and Enhances Children's Cognitive and Social-Emotional Growth*. New York: Oxford University Press.

Sklad, M., Diekstra, R., Ritter, M. D., Ben, J. & Gravesteijn, C. (2012). Effectiveness of school based universal social, emotional, and behavioral programs: Do they enhance students' development in the area of skill, behavior, and adjustment? *Psychology in the Schools*, *49*(9), 892-909.

Smetana, J. G. (1981). Preschool children's conceptions of moral and social rules. *Child Development*, *52*(4), 1333-1336.

Smetana, J. G. (2006). Social-cognitive domain theory: Consistencies and variations in children's moral and social judgments. In M. Killen & J. Smetana (Eds.), *Handbook of moral development* (pp. 119-153). New York: Taylor & Francis.

Smetana, J. G., Campione-Barr, N. & Yell, N. (2003). Children's moral and affective judgments regarding provocation and retaliation. *Merrill-Palmer Quarterly*, *49*(2), 209-236.

Smetana, J. G. & Killen, M. (2008). Moral cognition, emotions, and neuroscience: An integrative developmental view. *International Journal of Developmental Science, 2*(3), 324-339.

Smith, P. K., Ananiadou, K. & Cowie, H. (2003). Interventions to reduce school bullying. *The Canadian Journal of Psychiatry*, *48*(9), 591-599.

Smith, J. D., Cousins, J. B. & Stewart, R. (2005). Antibullying interventions in schools: Ingredients of effective programs. *Canadian Journal of Education/Revue canadienne de l'éducation*, 739-762.

Smith, J. D., Schneider, B. H., Smith, P. K. & Ananiadou, K. (2004). The effectiveness of whole-school antibullying programs: A synthesis of evaluation research. *School Psychology Review, 33*(4), 547.

Sneed, C. D. (2002). Correlates and implications for agreeableness in children. *The Journal of Psychology*, *136*(1), 59-67.

Snijders, T.A.B. & Bosker, R.J. (2011). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling.* New York: Sage Publishers.

Soepboer, S., Veenstra, R. & Verhulst, F. C. (2006). *Ben ik te min? Aard, omvang en achtergronden van antisociaal gedrag aan het einde van de basisschool en het begin van het voortgezet onderwijs*. Den Haag: WODC, Ministerie van Justitie.

Srivastava, S., John, O. P., Gosling, S. D. & Potter, J. (2003). Development of personality in early and middle adulthood: Set like plaster or persistent change?. *Journal of Personality and Social Psychology*, *84*(5), 1041.

Stavrinides, P., Georgiou, S. & Theofanous, V. (2010). Bullying and empathy: a short-term longitudinal investigation. *Educational Psychology*, *30*(7), 793-802.

Steiger, J. H. (1998). A note on multiple sample extensions of the RMSEA fit index. *Structural Equation Modeling: A Multidisciplinary Journal*, *5*(4), 411-419.

Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., ... & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *British Medical Journal*, *338*, b2393.

Stormshak, E. A. & Webster-Stratton, C. (1999). The qualitative interactions of children with conduct problems and their peers: Differential correlates with self-report measures, home behavior, and school behavior problems. *Journal of Applied Developmental Psychology*, *20*(2), 295-317.

Supovitz, J. A. & Weinbaum, E. H. (2008). *The Implementation Gap*. New York, NY: Teachers College.

Sutton, J., Smith, P. K. & Swettenham, J. (1999). Social cognition and bullying: Social inadequacy or skilled manipulation?. *British Journal of Developmental Psychology*, *17*(3), 435-450.

Tackett, J. L. (2006). Evaluating models of the personality–psychopathology relationship in children and adolescents. *Clinical Psychology Review*, *26*(5), 584-599.

Tangney, J.P. & Dearing, R.L. (2002). *Shame and Guilt*. New York: Guilford Press.

Tangney, J. P., Stuewig, J. & Mashek, D. J. (2007). Moral emotions and moral behavior. *Annual Review of Psychology*, *58*, 345.

Tani, F., Greenman, P. S., Schneider, B. H. & Fregoso, M. (2003). Bullying and the big five a study of childhood personality and participant roles in bullying incidents. *School Psychology International*, *24*(2), 131-146.

Tappan, M. B. (2006). Moral functioning as mediated action. *Journal of Moral Education*, *35*(1), 1-18.

Thompson, E. R. (2008). Development and validation of an international English big-five mini-markers. *Personality and Individual Differences*, *45*(6), 542-548.

Thornberg, R. (2008). School children's reasoning about school rules. *Research Papers In Education*, *23*(1), 37-52.

Tippett, N. & Wolke, D. (2014). Socioeconomic status and bullying: a meta-analysis. *American Journal of Public Health*, *104*(6), 48-59.

Tobin, R. M., Graziano, W. G., Vanman, E. J. & Tassinary, L. G. (2000). Personality, emotional experience, and efforts to control emotions. *Journal of Personality and Social Psychology*, *79*(4), 656.

Tucker, L. R. & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*(1), 1-10.

Turiel, E. (1983). *The Development of Social Knowledge: Morality and Convention.* Cambridge: Cambridge University Press.

Turiel, E. (1998). *The development of morality.* In W. Damon (Series Ed.) & N. Eisenberg (Volume Ed.), *Handbook of Child Psychology: Vol. 3 Social, Emotional, and Personality Development* (pp. 863-932). New York NY: John Wiley & Sons, Inc.

Turiel, E. (2006). Thought, emotions, and social interactional processes in moral development. In M. Killen & J. Smetana (Eds.), *Handbook of Moral Development* (pp. 7-35). New York: Taylor & Francis.

United Nations Educational, Scientific, and Cultural Organization (2011). *International Standard Classification of Education 2011.* Montreal, Canada: UNESCO Institute for Statistics.

Urbain, E. S. & Kendall, P. C. (1980). Review of social–cognitive problem-solving interventions with children. *Psychological Bulletin*, *88*(1), 109.

Van Noorden, T. H., Haselager, G. J., Cillessen, A. H. & Bukowski, W. M. (2015). Empathy and involvement in bullying in children and adolescents: A systematic review. *Journal of Youth and Adolescence, 44*(3), 637-657.

Veenstra, R., Dijkstra, J. K. & Kreager, D. A. (in press). Social development: Pathways, networks, and norms. In W. M. Bukowski, B. Laursen, & K. H. Rubin (Eds.), *Handbook of Peer Interactions, Relationships, and Groups* (pp. 45-63). New York: Guilford.

Veenstra, R., Lindenberg, S., Oldehinkel, A. J., De Winter, A. F., Verhulst, F. C. & Ormel, J. (2005). Bullying and victimization in elementary schools: a comparison of bullies, victims, bully/victims, and uninvolved preadolescents. *Developmental Psychology*, *41*(4), 672.

Velicer, W. F. & Fava, J. L. (1998). Affects of variable and subject sampling on factor pattern recovery. *Psychological Methods, 3*(2), 231-251.

Vollmeyer, R. Jenderek, K. &, Tozman, T. (2013). How different motivational aspects can affect moral behaviour. In K. Heinrichs, F. Oser &T. Lovat (Eds.), *Handbook of Moral Motivation* (pp. 267-287). Rotterdam: SensePublishers.

Von Rueden, U., Gosch, A., Rajmil, L., Bisegger, C. & Ravens-Sieberer, U. (2006). Socioeconomic determinants of health related quality of life in childhood and adolescence: results from a European study. *Journal of Epidemiology & Community Health, 60*(2), 130-135.

Vozzola, E.C. (2014). *Moral Development: Theory and Application.* New York: Routledge, Taylor & Francis Group.

Walker, L. J. (1999). The perceived personality of moral exemplars. *Journal of Moral Education, 28*(2), 145-162.

Walker, L. J. & Hennig, K. H. (2004). Differing conceptions of moral exemplarity: just, brave, and caring. *Journal of Personality and Social Psychology, 86*(4), 629.

Warden, D., Cheyne, B., Christie, D., Fitzpatrick, H. & Reid, K. (2003). Assessing children's perceptions of prosocial and antisocial peer behaviour. *Educational Psychology, 23*(5), 547-567.

Warden, D. & Mackinnon, S. (2003). Prosocial children, bullies and victims: An investigation of their sociometric status, empathy and social problem-solving strategies. *British Journal of Developmental Psychology, 21*(3), 367-385.

Weare, K. & Nind, M. (2011). Mental health promotion and problem prevention in schools: what does the evidence say? *Health Promotion International, 26*(1), 29-69.

Weisberg, Y. J., DeYoung, C. G. & Hirsh, J. B. (2011). Gender differences in personality across the ten aspects of the Big Five. *Frontiers in Psychology, 2,* 178.

Whitney, I. & Smith, P. K. (1993). A survey of the nature and extent of bullying in junior/middle and secondary schools. *Educational research*, *35*(1), 3-25.

Wienke, D., Anthonijsz, I., Abrahamse, S., Daamen, W. & Nieuwboer, A. (2014). *Beoordeling anti-pestprogramma's: Rapportage van de commissie voor het Ministerie Onderwijs, Cultuur en Wetenschap (OCW).* Utrecht: Nederlands Jeugdinstituut.

Wiersma, N. & Laupa, M. (2000). Young children's conceptions of the emotional consequences of varied social events. *Merrill-Palmer Quarterly, 46*(2), 325-341.

Willems, F., Denessen, E., Hermans, C. & Vermeer, P. (2012). Students' Perceptions and Teachers' Self-Ratings of Modelling Civic Virtues: An Exploratory Empirical Study in Dutch Primary Schools. *Journal of Moral Education, 41*(1), 99-115.

Zahn-Waxler, C., Cole, P. M., Welsh, J. D. & Fox, N. A. (1995). Psychophysiological correlates of empathy and prosocial behaviors in preschool children with behavior problems. *Development and Psychopathology, 7*(01), 27-48.

Zhou, Q., Valiente, C. & Eisenberg, N. (2003). Empathy and its Measurement. In S. J.Lopez & C.R.Snyder (Eds.), *Positive Psychological Assessment: A Handbook of Models and Measures* (pp. 269–284). Washington, DC: American Psychological Association.

ICO Dissertation Series

In the ICO Dissertation Series the dissertations of graduate students from faculties and institutes on educational research within the ICO Partner Universities are published: Eindhoven University of Technology, Leiden University, Maastricht University, Open University of the Netherlands, University of Amsterdam, University of Twente, Utrecht University, VU University Amsterdam, and Wageningen University, and formerly University of Groningen (until 2006), Radboud University Nijmegen (until 2004), and Tilburg University (until 2002). The University of Groningen, University of Antwerp, University of Ghent, and the Erasmus University Rotterdam have been 'ICO 'Network partner' in 2010 and 2011. From 2012 onwards, these ICO Network partners are full ICO partners, and from that period their dissertations will be added to this dissertation series.

ICO Dissertations 2016:

326. Lucieer, S.M. (06-01-2016). *Selecting Students for Medical Education: Exploring Novel Approaches.* Rotterdam: Erasmus University Rotterdam.

327. Ouwehand, K.H.R. (15-01-2016). *Effects of Observing and Producing Deictic Gestures on Memory and Learning in Different Age Groups*. Rotterdam: Erasmus University Rotterdam.

328. Bogert, N.J. van der (21-01-2016). *On teachers' visual perception and interpretation of classroom events using eye tracking and collaborative tagging methodologies.* Eindhoven: Eindhoven University of Technology.

329. Wolff, C. (16-02-2016). *Revisiting 'withitness': Differences in teachers' representations, perceptions, and interpretations of classroom management.* Heerlen: Open University of the Netherlands.

330. Kok, E.M. (01-04-2016). *Developing visual expertise; from shades of grey to diagnostic reasoning in radiology.* Maastricht: Maastricht University.

331. De Beer, H.T. (11-05-2016). *Exploring Instantaneous Speed in Grade Five: A Design Research.* Eindhoven: Eindhoven University of Technology.

332. Ebbeler, J. (12-05-2016). I*mplementing data use in schools: effects on the professional development of educators and the role of school leaders in data teams.* Enschede: University of Twente.

333. Draaijer, S. (10-06-2016). *Supporting Teachers in Higher Education in Designing Test Items.* Amsterdam: Vrije Universiteit Amsterdam.

334. Bos, L.T. (15-06-2016). *Moving Beyond Words. Supporting Text Processing Using a Situation Model approach.* Amsterdam: Vrije Universiteit Amsterdam.

335. Vrugte, J. ter (16-06-2016). *Serious support for serious gaming.* Enschede: University of Twente.

336. Kock, Z.D.Q.P. (23-06-2016). *Toward physics education in agreement with the nature of science: Grade 9 electricity as a case.* Eindhoven: Eindhoven University of Technology.

337. Trinh Ba, T. (28-6-2016) *Development of a course on integrating ICT into inquiry-based science education.* Amsterdam: Vrije Universiteit Amsterdam.

338. Gerken, M. (29-06-2016). *How do employees learn at work ? Understanding informal learning from others in different workplaces.* Maastricht: Maastricht University.

339. Louws, M.L. (06-07-2016) *Professional learning: what teachers want to learn.* Leiden: Leiden University.

340. Geel, M.J.M. van, & Keuning T. (08-07-2016). *Implementation and Effects of a Schoolwide Data-Based Decision Making Intervention: a Large-Scale Study.* Enschede: University of Twente.

341. Bouwer, I.R., & Koster, M.P. (02-09-2016) *Bringing writing research into the classroom: The effectiveness of Tekster, a newly developed writing program for elementary students.* Utrecht: Utrecht University.

342. Reijners, P.B.G. (02-09-2016.) *Retrieval as a Cognitive and Metacognitive Study Technique to Learn from Expository Text.* Heerlen: Open University of the Netherlands.

343. Hubers, M.D. (08-09-2016). *Capacity building by data team members to sustain schools' data use.* Enschede: University of Twente.

344. Hsiao, Y.P. (23-09-2016). *Peer Support to Facilitate Knowledge Sharing on Complex Tasks.* Heerlen: Open University of the Netherlands.

345. Scheer, E.A. (23-09-2016). *Data-based decision making put to the test.* Enschede: University of Twente.

346. Bohle Carbonell, K. (28-9-2016). *May I ask you….? The influence of Individual, Dyadic, and Network Factors on the Emergence of Information in Exchange Teams.* Maastricht: Maastricht University.

347. Claessens, L.C.A. (30-09-2016). *Be on my side, I'll be on your side: Teachers' perceptions of teacher–student relationships.* Utrecht: Utrecht university.

348. Jansen in de Wal, J. (18-11-2016). S*econdary school teachers' motivation for professional learning.* Heerlen: Open University of the Netherlands.

349. Kock, W.D. de. (24-11-2016). *The effectiveness of hints during computer supported word problem solving.* Groningen: University of Groningen.

350. Oonk, C. (07-12-2016). *Learning and Teaching in the Regional Learning Environment: Enabling Students and Teachers to Cross Boundaries in Multi-Stakeholder Practices'.* Wageningen: Wageningen University.

351. Beckers, J. (09-12-2016). *With a little help from my e-portfolio; supporting students' self directed learning in senior vocational education.* Maastricht: Maastricht University.

352. Osagie, E.R. (14-12-2016) *Learning and Corporate Social Responsibility. A study on the role of the learning organization, individual competencies, goal orientation and the learning climate in the CSR adaptation process.* Wageningen: Wageningen University.

Dankwoord

In het laatste deel van dit proefschrift wil ik graag iedereen bedanken die betrokken was bij mijn promotietraject. Zonder jullie bijdrage, interesse en steun was het me niet gelukt om een proefschrift te schrijven.

Allereerst heb ik geluk gehad met mijn drie begeleiders Greetje van der Werf, Marie-Christine Opdenakker en Tina Malti. Greetje, je hebt me altijd het gevoel gegeven dat je geloofde in mijn kunnen. Dit begon al tijdens het schrijven van mijn Bachelorscriptie, zette zich voort bij het schrijven van het onderzoeksvoorstel voor mijn promotietraject en bleef tot het einde van mijn promotie. Mijn proefschrift was zonder jouw scherpe geest en vermogen om hoofdzaken van bijzaken te onderscheiden niet hetzelfde geweest. Je zette me herhaaldelijk aan het denken en dwong me op deze manier om mijn ideeën helder te formuleren, zowel op papier als mondeling. Marie-Christine, bij jou kon ik altijd binnenlopen. Wanneer ik iemand nodig had om mijn gedachten te ordenen bood je een luisterend oor. Bovendien hebben jouw oog voor detail en kennis over analysetechnieken mijn proefschrift keer op keer aangescherpt. Wanneer jij een tekst van mij had gelezen, kon ik er van op aan dat je niets had gemist. Dit maakte het zowel eenvoudig als prettig om op je te bouwen. Tina, your expertise on moral development and developmental psychology made you a great advisor with regard to the set-up of my project, instrumentation, coding procedures, and the theoretical background of my articles. Also, you were able to challenge my thoughts without making me feel unknowledgeable. In this way I trusted your judgment and suggestions, motivating me to make the best of my work. Moreover, despite the ocean between us you were always available when I needed your help or advice. I truly appreciate this in you.

Naast mijn begeleiders wil ik ook graag de leden van de beoordelingscommissie (Prof. dr. D. Brugman, Prof. dr. R.J. Bosker en Prof. dr. R. Veenstra) bedanken voor het zorgvuldig lezen van mijn proefschrift en het geven van waardevolle feedback.

Ook enkele directe collega's verdienen een woord van dank. In chronologische volgorde, Coby, Mayra, Henriëtte, Manon, Sabine, Edwin en Nilla.

Coby en Mayra, jullie maakten het voor mij makkelijker om te beginnen met mijn promotie, omdat jullie me een goed voorbeeld gaven. Henriëtte en Manon, het delen van een kantoor met jullie was een erg gezellige ervaring waarbij veel gelachen werd. Sabine, jouw humor en empathie voor al mijn sores zorgden ervoor dat ik graag naar kantoor kwam, vaak met een nieuw verhaal waar je altijd belangstelling voor had. Ook maakte je me, waarschijnlijk onbedoeld, herhaaldelijk duidelijk wat mijn sterke kanten waren. Edwin, jij bent de meest attente jongen die ik ken. Wat hebben wij veel lol gehad: van telefoonalfabet tot sjoelcompetitie. Zelfspot en relativering vierden hoogtij in 'ons' kantoor. Zonder jou was mijn promotie veel zwaarder en eenzamer geweest. Nilla, met jou beleefde ik het leukste jaar van mijn promotie. Samen in jouw auto op pad naar een kleine verbetering van de wereld. Jouw kwaliteit om mensen echt te zien en jouw no-nonsense mentaliteit waren zowel voor mij, als voor mijn promotietraject, van onschatbare waarde. Twee indirecte collega's verdienen tot slot ook een aantal mooie woorden, Jeanine en Anouk. Jeanine, you taught me that chaos is fun and joy is around every corner. I truly admire your persistence and your aim to do good. The world should be happy to have you. Anouk, jouw scherpe geest en nieuwsgierigheid lieten me zien wat een 'echte' wetenschapper zou moeten zijn. Ik heb veel plezier beleefd aan jouw overdenkingen en heb bewondering voor je standvastigheid.

Dan Bineke Kloosterman, Vera Mak, Herman van der Molen, Jannis Bartelds en Lony Emmen. Bineke, jij hebt je meerdere jaren vrijwillig ingezet voor mijn project. Dit getuigt van toewijding en een warm hart, bedankt hiervoor. Vera, Herman, Jannis en Lony, wat hebben jullie veel werk verzet met het afluisteren van de interviews. Emma Wilson, bedankt voor je inzet om me mooie, maar ook passende tekeningen te bezorgen. Ton Kemp, bedankt voor de goede theateradviezen en je interesse in mijn onderzoek. Ginette Meijer, jij stond aan mijn zij aan het begin van mijn promotietraject. Bedankt voor je positieve houding, perfectionisme en onvermoeibaarheid tijdens je afstudeeronderzoek. Anke Kersting, Anita van Overmeeren, Maxime Schipper, Renate van der Veen en Sjieuwke Dankert, het begeleiden van jullie afstudeeronderzoeken leverde me telkens nieuw enthousiasme voor mijn eigen onderzoek op. Ook Monique Dijks, Eva Compaan, Marjolijn van der Gronden, Tessa van der Aart, Lieneke ten Brinke, Doreanne Tigchelaar wil ik bedanken voor hun inzet voor mijn onderzoek en de nieuwe ideeën die hieruit voortvloeiden.

Alle testleiders die mee hebben geholpen met de dataverzameling verdienen ook een woord van dank: Sarah van Tuinen, Nynke Hoogterp, Evelien Hoogterp, Christina Meijer, Ella ter Sluis, Silvia van den Berg, Jessica Boer, Wim Alberts, Bianca Bakker, Linda Goes, Mariëtte Hingstman, Annelore Roemers, Doutsen Klaver, Grytsje Talsma, Christianne Westerhof, Joke van Dijken, Anouk de Vries, Inge Hilbrands, Rianne Hornstra, Marlies Klassen-Vermaat, Inge Wichgers en Jaap van den Berg. Hetzelfde geldt voor de studenten die tijd hebben gestoken in de dataverzameling: Laura Moraal, Anna van Dijk, Simone Beek, Susan de Graaf, Anne van der Wolf, Riëtte Krijgsheld, Bo Wildschut, Silvia Hongerkamp, Marianne van Dijk, Melany Talstra, Anne-Grietje Heslinga, Myrthe Postema, Nikki Steenkist, Henriet Renkema, Britt Sinnema, Nynke Kiers, Margriet van der Heijden, Anika de Vries, Amarins Mous, Debby ten Hove, Ellen Brinks, Lisa Sterk, Babette Almoes, Esther Stellingwerf, Marloes Bakker, Gabriela Medina, Lianne de Graaf, Nadine Nixdorf en Evelien Luitjens. Heel erg bedankt.

Een belangrijke bijdrage aan mijn onderzoek kwam van de deelnemende scholen, de Sint Paulusschool, de Sint Bonifatiusschool, de Nynke van Hichtum, De Tsjerne, De Vlieger, De Meander, de Peter Petersen, de Beijumkorf, De Wierde, de Klim-op en de Kleine Dollart. Door het prettige contact met directie, leerkrachten en leerlingen heb ik veel geleerd over onderzoek en onderwijs, maar vooral over mensen en over mezelf.

In het vervolg wil ik graag een aantal mensen buiten mijn werk bedanken. First my family in Canada. David, Angela, Morgan, and Baukje, you hosted me during my stay in Toronto and taught me what hospitality means. Thank you for all the fun day-to-day moments that distracted me from any negative thought. Leo and Maureen, you also deserve some words. You hoped I would find my feet again at home and I did. However, I also found them with you and your lovely family. Dan mijn spreekwoordelijke familie in Amsterdam, de Nationale DenkTank ’15. Bedankt voor alle plezier en alles wat ik over mezelf heb geleerd. Jullie optimisme en verstrekkende acceptatie van ideeën maken dat ik met een warm gevoel terugkijk op een periode zonder proefschrift. Mijn spreekwoordelijke familie in Appelscha, de Humanitaskampleiding, wil ik graag bedanken voor alle nuchterheid die in een week op me inwerkte en alle humor die daarmee gepaard ging.

Ook mijn collega-studenten in de Research Master wil ik bedanken. Marijke, alhoewel je relatief ver weg was, hield ik met jou het meeste contact. Bij jou gaat het me makkelijk af om mezelf te zijn en dit is een gave die niet veel mensen

bezitten. Ik waardeer je openheid, relativeringsvermogen en inlevingsvermogen enorm. Zonder de talloze mailtjes, sms'jes, appjes en telefoontjes tussen ons beide had ik me in deze wereld veel minder begrepen gevoeld. Nynke, jouw onbevooroordeelde en positieve kijk maakten dat ik graag een kopje thee met je dronk. Je creativiteit en enthousiasme zijn bovendien erg aanstekelijk. Ruta, your warmth and ability to chat about almost everything made you a pleasant person to be around. You not only showed me your beautiful country, but also that being vulnerable is not a weakness. Christien, jouw verwondering over simpele dingen en bemoedigende persoonlijkheid deden me altijd goed als ik je op kwam zoeken in je kantoor 'aan de overkant'. Selwyn, jouw andere kijk op de wereld en je onophoudelijke stroom aan verhalen brachten de nodige afleiding tijdens mijn promotie. Tomas, je overdaad aan ideeën én de uitvoering ervan maken dat ik je erg bewonder. Ik heb meer dan eens tegen anderen gezegd dat er meer mensen zouden moeten zijn zoals jij. Felicity, your kindness, quality to stand out, and ability to beautifully phrase everything are a true example. Charlotte, bedankt voor alle levensanalyses, gedeelde liefde voor de limerick en je zorgzaamheid. Tot slot een docent die ik in de Research Master leerde kennen. Wendy, je was officieel niet mijn begeleider, maar hebt me op emotioneel vlak meer dan eens een prettige spiegel voorgehouden.

Mijn collega-studenten die ik ontmoette in de Bachelor Pedagogische Wetenschappen verdienen ook een speciale plek. Anna, Nynke, Laura, Simone en Trynke, bedankt voor alle gezellige ochtenden, middagen en avonden, en de kijkjes in jullie, inmiddels, zo uiteenlopende levens. Anna, voor je gangmakers-mentaliteit, Nynke, voor je creativiteit en onafhankelijke blik, Laura, voor alle keren dat ik tegen je aan mocht zeuren, Simone, voor de kunst om tevreden te zijn en Trynke, voor jouw vele-ballen-in-de-lucht. Mijn promotieonderzoek werd lichter door het contact met jullie. Ook Melanie en Hester leerde ik kennen tijdens mijn studietijd. Melanie, bedankt voor jouw betrokkenheid bij zowel mij als de wereld en Hester, bedankt voor jouw oprechte interesse en veelheid aan verhalen.

Ook mijn middelbare school is vertegenwoordigd. Erzsi en Anneke, wat is het eigenlijk bijzonder dat we elkaar nog steeds opzoeken. Anneke, jouw no-nonsense- mentaliteit is aanstekelijk en erg prettig. En Erzsi, bescheiden en onbevooroordeeld sta je altijd voor me klaar. Je hebt voor mijn onderzoek zelfs een aantal interviews voor me afgeluisterd. Bedankt hiervoor.

Markus, bedankt dat je bij me wilt horen. Je geeft me rust en energie tegelijkertijd. Door jou vind ik mijn gewone leven een stuk minder gewoon en dagelijkse dingen een stuk minder dagelijks. Door jou werd ook mijn proefschrift meer bijzonder.

Alya, wat ben ik blij met jou als zusje. Sterk als een beer, letterlijk en figuurlijk. Ik weet dat je altijd achter me zult staan en dat is een erg fijn gevoel. Beppe, mam, heit, jullie waren er vanaf mijn begin. Beppe, sels fan in afstand fiel ik dast do grutsk op my bist. Bedankt foar dyn belutsenens en bewondering. Mam, heit, op jullie kan ik bouwen. Heit, mei dyn humor kin ik de hiele wrâld oan en dyn analytisch vermogen makket dy in prettige gesprekspartner. Mam, jij hebt werkelijk alles voor me over. Met het werk dat jij hebt verzet tijdens mijn promotie verdien je een grote prijs. Je zei eens dat ik een boom groei, met af en toe een tak de één of andere kant op, maar dat ik altijd weer terug kom bij de stam. Die stam heb ik van jou en heit meegekregen, als sociaal betrokken, onafhankelijke mensen. Bedankt voor jullie helpende handen, geduld voor al mijn - vaak zelf verzonnen - beren op de weg, en het stevige gevoel dat ik altijd een thuis heb.

About the Author

Dorinde Jansma was born on the 9th of May 1988 in Dokkum, the Netherlands. After obtaining her VWO diploma (cum laude) in 2006 at the Dockinga College in Dokkum, she started studying Pedagogical and Educational Sciences at the University of Groningen. She obtained her Bachelor degree (cum laude) in 2009 with specializations in General Pedagogy and Educational Sciences and her Master's degree (cum laude) in 2010 with specializations in Educational Policy and Organization and Educational Design. She graduated (cum laude) from the Research Master Social and Behavioural Sciences in 2012 focusing on the social outcomes of education. Her engagement in statistics and research was not only reflected by the additional courses she took, but also by assisting in research, participating in the onset of an international research project, working in the Methodology shop, working as a thesis supervisor and teacher at the pre-Master Educational Sciences of SPO, and by teaching a variety of research and statistics courses in the Bachelor Pedagogical and Educational Sciences and Academic Teacher Training Education of the University of Groningen. She was also a member of the Groningen Talent Group, selected as one of the hundred most talented persons for the city that were between 20 and 30 years old.

During her PhD, that was funded by the Research Talent Grant of NWO, Dorinde was a visiting scholar working with Tina Malti at the University of Toronto. Also, she worked as a teacher and supervisor of students in the Bachelor Pedagogy and Educational Sciences, Master Educational Sciences, Research Master Social and Behavioural Sciences at the University of Groningen and in the pre-Master Educational Sciences of SPO. As part of her PhD she initiated and developed a program for primary education teachers to foster social-emotional development in primary school children. Her social involvement is also illustrated by her work with children with behavioural problems, volunteering jobs, membership of educational committees and of the KNAW Evaluation Committee Research Masters Behavioural Sciences. During her PhD she participated in the National Think Tank of 2015 in which twentyfour promising master- and PhD-candidates voluntarily spend four months to solve the complex societal issue of 'learning in the future'.

Dorinde is currently working as a University Lecturer specialized in Educational and Developmental Psychology at the University of Groningen.